Three Perspectives on Multilevel Selection:

An Experimental, Historical, and Synthetic Analysis of Group-Level Selection

by

Christopher C. Dimond

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved August 2014 by the
Graduate Supervisory Committee:

James P. Collins, Chair
Juergen Gadau
Manfred Laubichler
Brad Armendt
John Lynch

ARIZONA STATE UNIVERSITY

December 2014

ABSTRACT

During the 1960s, the long-standing idea that traits or behaviors could be explained by natural selection acting on traits that persisted "for the good of the group" prompted a series of debates about group-level selection and the effectiveness with which natural selection could act at or across multiple levels of biological organization. For some this topic remains contentious, while others consider the debate settled, even while disagreeing about when and how resolution occurred, raising the question: "Why have these debates continued?"

Here I explore the biology, history, and philosophy of the possibility of natural selection operating at levels of biological organization other than the organism by focusing on debates about group-level selection that have occurred since the 1960s. In particular, I use experimental, historical, and synthetic methods to review how the debates have changed, and whether different uses of the same words and concepts can lead to different interpretations of the same experimental data.

I begin with the results of a group-selection experiment I conducted using the parasitoid wasp *Nasonia*, and discuss how the interpretation depends on how one conceives of and defines a "group." Then I review the history of the group selection controversy and argue that this history is best interpreted as multiple, interrelated debates rather than a single continuous debate. Furthermore, I show how the aspects of these debates that have changed the most are related to theoretical content and empirical data, while disputes related to methods remain largely unchanged. Synthesizing this material, I distinguish four different "approaches" to the study of multilevel selection based on the questions and methods used by researchers, and I use the results of the *Nasonia*

experiment to discuss how each approach can lead to different interpretations of the same experimental data. I argue that this realization can help to explain why debates about group and multilevel selection have persisted for nearly sixty years. Finally, the conclusions of this dissertation apply beyond evolutionary biology by providing an illustration of how key concepts can change over time, and how failing to appreciate this fact can lead to ongoing controversy within a scientific field.

ACKNOWLEDGMENTS

This work was supported by an NSF Graduate Research Fellowship.

TABLE OF CONTENTS

Page

LIST OF TABLES vi

LIST OF FIGURES vii

CHAPTER

INTRODUCTION 1

1 A TEST OF GROUP-LEVEL SELECTION FOR INCREASED POPULATION SIZE IN THE PARASITOID WASP *NASONIA VITRIPENNIS* 5

Abstract 5

Introduction 6

Methods 7

Results 14

Discussion 17

Conclusion 25

2 WHY ARE WE STILL ARGUING ABOUT GROUP SELECTION? 28

Abstract 28

Introduction 29

Group Selection Before 1983 31

Current Group Selection Debates and Multilevel Selection 48

Multile Histories 74

Conclusion 76

CHAPTER Page

3 DIFFERENT WAYS OF STUDYING MULTILEVEL SELECTION AND THEIR IMPLICATIONS FOR EXPERIMENTAL DATA AND THEORETICAL ARGUMENTS 79

Abstract 79

Introduction 79

Characterizing Research on Multilevel Selection 81

Approaches to Multilevel Selection 85

Interpretation of Experimental Data 94

Conflicting Interpretation 99

Conclusion 109

CONCLUSION 112

REFERENCES 116

APPENDIX

A SUPPLEMENTAL MATERIALS FOR CHAPTER 1 128

LIST OF TABLES

Table		Page
1.1	ANOVA Table 1.1	15
1.2	ANOVA Table 1.2	15
1.3	ANOVA Table 1.3	15
1.4	ANOVA Table 1.4	16
1.5	Definitions of "Group"	21
2.1	Experimental Studies of Group and Multilevel Seleciton	60
2.2	Group Selection Criticisms Addressed by Eldakar and Wilson (2011)	62

LIST OF FIGURES

Figure		Page
1.1	Comparison of Organizational Levels for *Nasonia* Habitat	9
1.2	Full Experimental Design	12
1.3	Population Size Results	16
1.4	Sex Ratio Results	17
3.1	Representation of Approahes to the Study of Multilevel Selection	82
S.1	Creation of the ASURx Laboratory Population	130

INTRODUCTION

Few subjects in biology are as fundamental to the discipline as the study of natural selection. Arguments about the function and effectiveness of natural selection date to Darwin's work before publication of the first edition of *On The Origin of Species* in 1859 and remain a consistent part of discussions about evolutionary theory. Although the theory of evolution by natural selection is no longer a controversial topic in evolutionary biology, the question of how or if natural selection works at levels of organization other than the individual remains contentious.

The idea of "group selection" or natural selection acting upon groups of organisms rather than only on organisms or their genes has a long history of use, by some readings dating back to Darwin's own work (Borrello 2010). A series of rigorous discussions during the 1960s about whether group-level selection could be an effective cause of an evolutionary response triggered both controversy and debates about the topic and led to decades of arguments and disagreements about the appropriate way(s) to conceptualize selection working across multiple levels of biological organization.

My project concerns why controversy has continued for one particular case of multilevel selection regarding groups as a level at which natural selection may or may not operate effectively. In particular, my research affords a way to understand the contemporary group-selection debate in its historical, scientific, and philosophical contexts by synthesizing across these disciplines. In doing so, I illustrate the role that conceptual change has played in the controversy about group-level selection over the past sixty years, and argue that both conceptual clarity as well as empirical evidence are

necessary to resolve the controversy. While the necessity of conceptual clarity might seem obvious, and some researchers worked to create such clarity, my research demonstrates that the failure by some to acknowledge how concepts change over time has contributed to the ongoing controversy within this field of study. Finally, understanding the conceptual change within this controversy helps us to appreciate the current state of knowledge regarding multilevel selection and reveals the varied definitions, questions, and methods that future researchers will want to be aware of for their own projects.

The first chapter of this dissertation describes an experiment that asked whether groups of five female wasps housed collectively would respond to selection for increased population size. Yet, as I describe, interpretation of the otherwise straightforward results is confounded by a multiplicity of implicit definitions for the term "group." In other words, despite decades of theoretical and empirical research on the topic of group-level selection, interpretations of empirical data vary depending upon the definition of "group" that is used.

The second chapter builds upon this conclusion by asking why the controversy surrounding the study of group-level selection has continued for so long, particularly highlighting the fact that researchers disagree about if and when any resolution occurred. In this chapter, I compare the group selection controversy of the 1980s, represented by David Sloan Wilson's (1983) review, with the current state of discussions and debates regarding both group-level selection and the broader theoretical framework of multilevel selection (MLS). In doing so, I provide a review of modern MLS theory as well as demonstrate the ways in which the debates about levels of selection have and have not changed. Using this account, I argue that the history of the group selection controversy is

best interpreted as multiple, interrelated debates rather than a single continuous debate. Furthermore, I show the aspects of these debates that have changed the most are those related to the theoretical content and empirical data available, while disputes related to methods remain largely unchanged.

Finally, the third chapter builds upon the first two in order to provide a framework for understanding the ongoing disputes about group selection, and multilevel selection more generally, by arguing that there are different "approaches" to the study of multilevel selection distinguished by where the research falls upon two axes: 1) whether the researcher is primarily interested in a single level of organization or multiple levels, and 2) the extent to which the work starts from a theoretical or empirical perspective. I then describe four such "approaches" including the questions asked and the methods used and discuss how these approaches might lead to conflicting interpretations of empirical data by using the *Nasonia* experiment from chapter 1 as an example. By acknowledging these different approaches, it becomes clear why some of the controversy surrounding group selection continues, and in doing so provides potential avenues for avoiding the same conflicts in future research.

This project offers valuable insights for other fields of study that rely upon an understanding of natural selection. In particular, researchers in the burgeoning study of complex adaptive systems (CASs)—hierarchically organized systems in which lower level entities interact to produce emergent properties at higher levels of organization—have drawn upon evolutionary theory and natural selection, either implicitly or explicitly, when analyzing CASs since one key element of a CAS is its ability to respond to various selection pressures, which is easily seen in biological CASs.

This CAS interest in selection working across hierarchal levels has created a strong theoretical dependence not only upon biologists' notions of natural selection, but natural selection occurring at multiple levels of organization, either by considering what kinds of levels or organizations are capable of responding to selection or by considering how selection might act simultaneously across levels. Because of this, it is particularly important to understand the ways that biologists have conceptualized natural selection occurring across multiple levels of biological organization. Of course, the ongoing controversy on the subject can make this a challenging task. But this project offers an account of how biologists and philosophers of biology have conceptualized natural selection working across multiple levels of biological organization over the last 60 years.

Finally, the lessons that this project offers extend beyond the subject of natural selection. As David Sloan Wilson (1983) noted, a description of the history, philosophy, and biology involved in the controversy over group selection provides a striking example of scientific discourse, including a demonstration of the interrelated roles of empirical evidence and theory generation, as well as an illustration of the benefits that philosophers of biology can offer to subjects such as this. But more than that, this dissertation also illustrates how the use of key concepts can change over time, and how failing to acknowledge and appreciate this fact can lead to ongoing controversy within a scientific field of study. The conclusion that there are multiple approaches to the study of multilevel selection also offers insights for other sciences by demonstrating how different interpretations of empirical data can arise and suggesting a way of understanding ongoing controversy in other fields of research.

CHAPTER 1

A TEST OF GROUP-LEVEL SELECTION FOR INCREASED POPULATION SIZE IN THE PARASITOID WASP *NASONIA VITRIPENNIS*

Abstract

I describe the results of a selection experiment using the parasitoid wasp *Nasonia* to test if five females housed in a single vial would respond to selection for increased population size at the "group-level" collective of five wasps. I show no response to selection in the form of an increase in mean population size; however, the sex ratio of the group treatment did change. The results of this experiment also suggest that female *Nasonia* adjust the sex ratio of their broods in ways that at least partially depend upon conditions that they have experienced previously, by increasing the relative proportion of males in future broods after they have encountered other females, or decreasing the proportion of males in future broods after initially encountering no other females.

In discussing the outcome of this experiment, I highlight the challenge of interpreting the results because any interpretation is predicated upon how one conceives of a "group." That is, by some definitions of "group" this experiment was a case of group-level selection that failed to result in a response to the target of selection, while by other definitions of "group" this experiment failed to include group-level selection at all. That such varied interpretations arise from different definitions of a group demonstrates the importance of understanding the history of such conceptual change, and why experimentalists must be clear about the concepts upon which their research designs are based since conceptual change might influence the interpretation of their results.

Introduction

The possibility that natural selection might act at more than one level of biological organization—either independently or simultaneously—has a rich history of theoretical analysis (Borello 2010; Hamilton and Dimond 2012). There are also empirical analyses examining the levels at which selection might act that cover a range of experimental conditions and study organisms. These have included flour beetles (Wade 1976, 1977, Craig, 1982; Wade and Goodnight, 1991), cress and jewelweed plants (Goodnight 1985; Stevens *et al*. 1995), domesticated chickens (Muir 1996, Craig and Muir 1996), livebearing fish (Baer *et al*. 2000), parasitic nematodes (Bashey and Lively 2009), as well as other species summarized by Goodnight and Stevens (1997).

Of the above studies, only one used a parasitic species, and none of the studies used a haplodiploid species. Continuing to add to the species and conditions studied at the group level will add to our comparative understanding of the possibility of selection at multiple levels. Here, I present the results of a selection experiment with the parasitoid wasp *Nasonia vitripennis*. *Nasonia* has a long history as a study system for facultative sex ratios and the effects of inclusive fitness (Werren 1980; Parker and Orzack 1985; Pannebakker *et al*. 2011). Additionally, *Nasonia* behavior, including the ability of females to adjust the sex ratio of their broods (Burton-Chellew et al, 2008), and the patchy distribution of their hosts (Grillenberger *et al*. 2008), suggest that *Nasonia* might be subject to selection at levels other than just the organism.

I asked if a collective of five female *Nasonia* housed in a single vial (patch) would respond to selection for increased population size and compared this response to

the response when females were housed alone and selected for increase number of offspring. Thus, the targets of selection for this multilevel selection experiment were the "group-level" collective of five wasps, following similar designs in other studies of group-level selection (e.g., Wade 1976, 1977), and the organismal-level individuals.

I predicted that there would be a response to selection in the group-level treatment because of the similarities between my experimental design and other group-level selection experiments (e.g., Wade 1976; Bashey and Lively 2009). Specifically, I expected to see an increase in mean population size. Furthermore, because of previous research on the ability of *Nasonia* to change sex ratios facultatively for each clutch of eggs, I also predicted that the sex ratios would show an increase in the number of males relative to females as part of the overall response to selection on groups of wasps.

Methods

Study organism: The parasitoid wasp genus *Nasonia* has four species: *N*. *vitripennis*, *N*. *giraulti*, *N*. *longicornis*, and *N*. *oneida*. Three of the four species cannot cross-fertilize due to cytoplasmic incompatibilities caused by their endosymbiont *Wolbachia* (Bordenstein and Werren 2007). *Nasonia* have a haplodiploid sex determination system and, because *Nasonia* generation times are approximately 14 days at 24 degrees C, have been used as a study system to test sex ratio theory (Werren 1980; Parker and Orzack 1985; Orzack et al 1991; Ivens *et al*. 2009; Beukeboom and Van de Zande 2010). Knowledge of the complete genome of *N*. *vitripennis* (Werren *et al*. 2010) also opens the possibility for researchers to track genetic changes underlying phenotypic variation. Although popular as an evolutionary genetics study system (Beukeboom and

Desplan 2003), *Nasonia* species have not been used for exploring questions related to levels of selection even though the same characteristics that make *Nasonia* an ideal system for experimental studies of evolution and genetics also suggest that it could be useful for multilevel selection studies.

Experimental population: I used the ASU Remix (ASURx) population, a laboratory population of *N*. *vitripennis* created for this project by combining eight *N*. *vitripennis* laboratory strains (see Appendix A).

Selection target: Numerous experiments studying group and multilevel selection have used the trait "population size" as the target of selection (e.g., Wade 1976, 1977; Wade and Goodnight 1990a, 1990b; Baer *et al*. 2000; Bashey and Lively 2009). I used the same trait for this experiment in order to compare my results with previous studies. However, because *Nasonia* is a parasitic wasp, the available resources – fly larvae – had to be standardized. I did this by selecting upon the number of adult wasps per standard resource, as described below. In each treatment I maintained a larval host-to-wasp ratio of 2:1, and I used the total mass of hosts to standardize the available resources for each group. I then counted the number of adult wasps that emerged in each vial and divided by number of hosts successfully parasitized, so as not to penalize groups for any dead hosts. Thus, the direct target of selection was number of adult offspring per standardized host. Defining and selecting upon population size in this way also allowed for the same character to be selected in the group and organismal treatments, as further described in the experimental design section.

Group size: *N*. *vitripennis* females parasitize blowfly and fleshfly pupae found in bird nests or dung piles (Fig. 1.1), typically laying between 20 and 40 eggs in a single

host (Werren, 1980). Multiple females often parasitize a single patch, and the mean number of *N*. *vitripennis* foundresses vary from 2.3 per nest (limits 1 to 9) (Molbo and Parker, 1996) to 3 per patch (limits 1 to 7) (Grillenberger *et al*. 2008). Multiple females will also parasitize the same host pupa, with observed rates of superparasatism—parasitism of a host by more than one female *Nasonia*—ranging from 40% of all sampled patches containing a superparasatized host (Grillenberger *et al*. 2008) to 100% of the patches visited by more than one female containing a superparasatized host (Molbo and Parker, 1996).

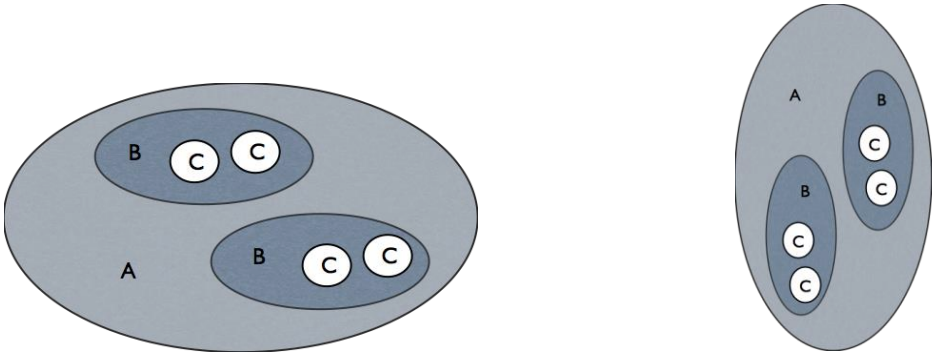


Figure 1.1 Comparison of organizational levels for *Nasonia* habitat in nature and laboratory.
Image on the left represents the levels of organization in a bird nest (A) with fly pupae (B) within which *Nasonia* eggs develop (C). Image on the right represents levels of organization in the laboratory, with vials (A) containing the fly pupae (B) in which *Nasonia* eggs develop (C).

To approximate conditions similar to those observed, I created "groups" for this experiment by placing five mated females into a single vial with access to 10 fly pupae as hosts (Fig 1.1). While this number of wasps is greater than the mean number that would typically parasitize a single patch, it is within the limits of observed populations. Creating groups with five females also increased the likelihood of superparasitism and other forms of interaction between individuals within each group, which is essential for some

definitions of a "group" (e.g., Wilson, 1975) and is thought to be important for increasing the potential for between-group differences upon which selection can act (e.g., Goodnight and Stevens 1997).

Experimental design: I used a nested design with two treatments: the group treatment and the organismal treatment. Within each treatment were two environmental conditions that each female encountered: a collective context with a group of five wasps and a solitary context with each wasp housed alone. The entire experiment had three replicates.

The purpose of the group treatment was to test if groups of five female wasps in a single vial with equal access to a common store of hosts would respond to selection on the size of the population in the vial. After 2 days, these same females were transferred individually into the solitary context vials, maintaining the same ratio of hosts per wasp. This solitary context following the collective context provided a comparison of the females' behavior when not in the presence of other females, including an analysis of how the mean population size produced by the group compared with the total offspring of the same wasps housed separately. Note that within the group treatment, artificial selection was only exerted upon the collective context.

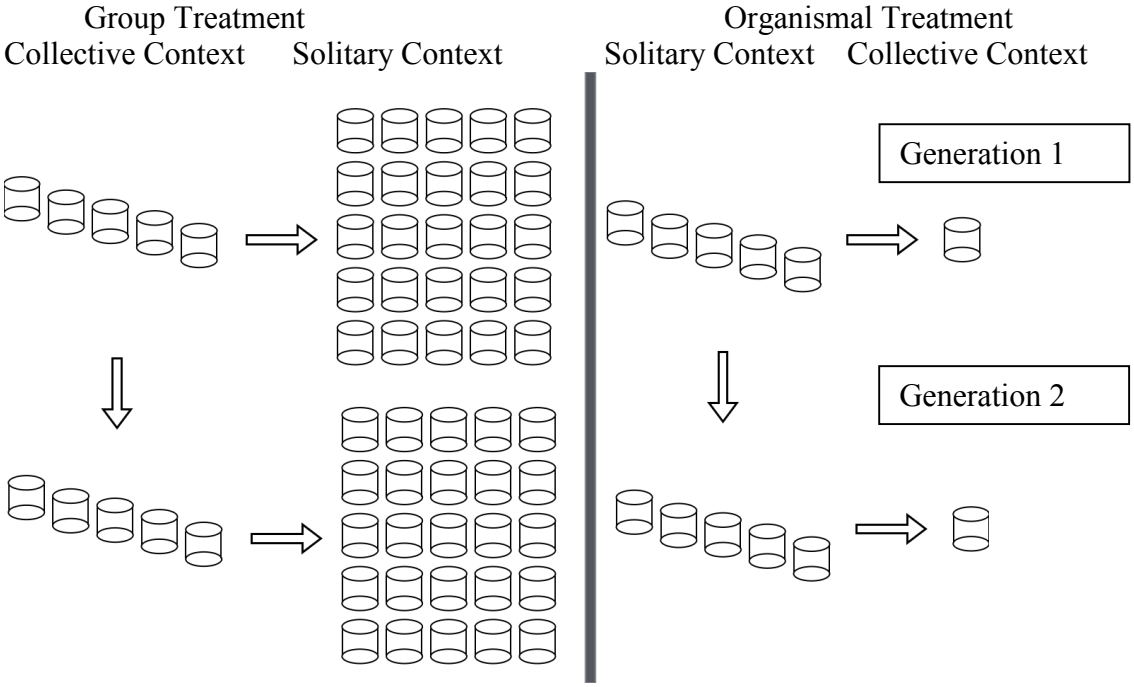
The purpose of the organismal treatment was to serve as a comparison to the group treatment by selecting on the same character in a solitary context. Because the comparison was intended to be at the level of the vial—for population sizes—rather than match the same number of females between the group and organismal treatments, I matched the number of vials. Thus, I had 5 group vials in each replicate and 5 organismal vials in each replicate (Figure 1.2). In the organismal treatment, the collective context,

with all 5 wasps transferred to a single vial, served as a comparison for the solitary context and also provided a comparison to the collective context of the group treatment.

Selection protocol: For the group treatment selection only occurred on the collective context vials. For this context, I placed 5 mated females into each of 5 vials for a total of 25 females per replicate (Fig. 1.2a). I then added 10 larval hosts to each vial, creating the wasp-to-host ratio of 1:2. The females were allowed approximately 48 hours to lay on the hosts and then each wasp was transferred to a separate vial with only 2 hosts, maintaining the 1:2 wasp to host ratio (Fig. 1.2b) for the solitary context. The wasps were then allowed approximately 48 hours to oviposit in the hosts.

Fourteen days after the wasps in each treatment were given hosts, I counted the number of wasps that had emerged from the host puparia and standardized this count against the available resources for each group. I did so by dividing the total number of wasps by the original mass (g) of the hosts placed in that collective vial. Because some hosts were not parasitized, I then adjusted the standardized count by dividing by the total number of hosts that were parasitized, including those from which adult wasps emerged and those from which no adults emerged but within which were developing *Nasonia* pupae. This method provided the number of adult wasps per standardized host for each collective vial, which I used as the basis for selecting the vial with the largest population size in each replicate. I used the vial chosen to create 5 new vials with 5 mated females and 10 hosts each to start the next generation. In other words, this experiment used a propagule pool as opposed to a migrant pool (Wade 1978), as the former is more likely to favor group-level selection. I repeated this process over 5 rounds of selection creating Generations 0, 1, 2, 3, 4, and 5.

1.2a. Full Experimental Design



1.2b. Transfer of wasps between contexts

1.2c. Creation of next generation

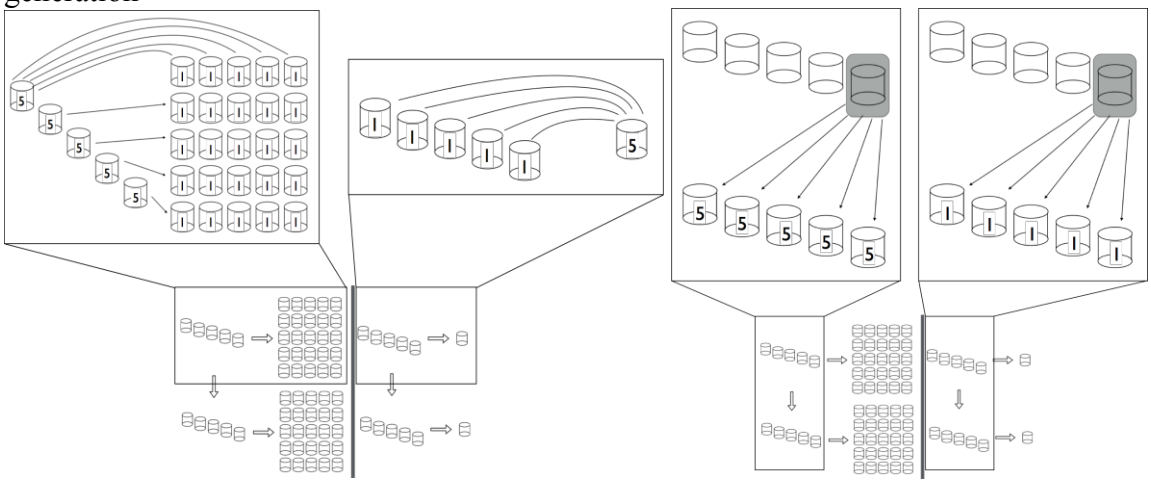


Figure 1.2. Experimental Design. In each section, the group treatment is on the left and the organismal treatment on the right. 1.2a) Depiction of the full experimental design, shown representing two of the total generations. 1.2b) Detailed representation of the horizontal arrows depicting the transfer of contexts shown in 1.2a. 1.2c) Detailed representation of the vertical arrows depicting the selection and creation of subsequent generations shown in 1.2a.

The protocol for the organismal treatment was identical, with the exception that selection only occurred on the solitary context and each replicate started with 5 females in the solitary context and these were moved to the collective context (1.2b). The timing between these transfers remained the same. After assessing the total offspring for each solitary female, I used the same calculations as described above—including host weights and the number parasitized—and then chose the brood with the most adults per standardized host in each replicate. From this brood I collected 5 mated females to begin the next generation of the organismal selection treatment by placing each mated female into a separate vial with 2 hosts (Fig. 1.2c).

Analysis: Because the goal of this experiment was to estimate the overall response or lack of response to selection, I used an analysis comparing the number of offspring per standardized host in the first generation of the experiment and in the last generation of the experiment in A) each treatment: group and organismal, and in B) each context: collective and solitary. For these comparisons, I used one-way ANOVAs since the observations were approximately normally distributed and did not show significant differences in variance.

I also tested if sex ratios changed during the experiment, again comparing the first and last generations. For these analyses I used 2-sample t-tests with the conservative assumption of unequal variances in the samples and used a 2-tail test for differences above or below the mean.

To estimate organismal contributions to group phenotypes, I designed this experiment to use contextual analysis, a technique from the social sciences (e.g., Boyd and Iversen 1979), but also used in biology over the last 20 years (Heisler and Damuth

1987; Goodnight *et al*. 1992; Okasha 2006). Contextual analysis requires knowing the organismal contribution to group fitness, which is unavailable for this experiment. This is why I added the solitary context to the group treatment to serve as a proxy for organismal fitness and relative contribution to the population. However, because of the significant differences in the behavior of females when solitary compared with when they were in groups, particularly with regard to number of eggs laid and sex ratio of the broods (see results section), the solitary context would not have served as a reliable proxy for the organismal contributions needed for contextual analysis. As such, I decided that contextual analysis would have been inappropriate and removed it from the final analysis.

Results

Population size: In the collective vials of five wasps within the group treatment, there was no statistically significant difference between mean number of offspring per host in Generation 0 and mean number of offspring per host in Generation 5 (Table 1.1). However, when the same females were later allowed to lay eggs in the solitary context of the group treatment, there was a significant increase in mean number of offspring per host between Generations 0 and 5 (Table 1.2). In the organismal treatment, there was a significant increase in mean number of offspring per host when the females were housed individually (Table 1.3). Similar to the group treatment, in the organismal treatment there was no significant change between Generation 0 and Generation 5 of the experiment when females were housed collectively in groups of five after they were hosted alone (Table 1.4). Furthermore, in both selection treatments, females laying in solitary contexts produced a mean number of offspring per host that was significantly greater (six fold difference) than females allowed to lay collectively in group contexts (Figure 1.3).

Table 1.1. ANOVA table for comparison of beginning and ending mean number of offspring per host in the group treatment for collective vials containing five females

SUMMARY

Groups	*Vials*	*Sum*	*Mean*	*SEM*
Gen 0	15	311	20.74	1.82
Gen 5	15	287	19.12	1.14

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	19.56	1	19.56	0.567	0.46	4.2
Within Groups	966.50	28	34.52			
Total	986.10	29				

Table 1.2. ANOVA table for comparison of beginning and ending mean number of offspring per host in the group treatment for females in solitary vials

SUMMARY

Groups	*Vials*	*Sum*	*Mean*	*SEM*
Gen 0	55	7190	131	7.99
Gen 5	55	8783	160	7.34

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	23069	1	23069	7.6108	0.007	3.93
Within Groups	327357	108	3031			
Total	350426	109				

Table 1.3. ANOVA table for comparison of beginning and ending mean number of offspring per host in the organismal treatment for females in solitary vials

SUMMARY

Groups	*Vials*	*Sum*	*Mean*	*SEM*
Gen 1	12	1586	132	16.4
Gen 5	12	2265	189	13.75

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	19236	1	19236	7.00	0.015	4.30
Within Groups	60488	22	2749			
Total	79724	23				

Table 1.4. ANOVA table for comparison of beginning and ending mean number of offspring per host in the organismal treatment for collective vials containing five females

SUMMARY

Groups	*Vials*	*Sum*	*Mean*	*SEM*
Gen 1	3	85.72	28.57	1.63
Gen 5	3	114.9	38.30	3.82

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	141.78	1	141.78	5.48	0.08	7.71
Within Groups	103.44	4	25.86			
Total	245.22	5				

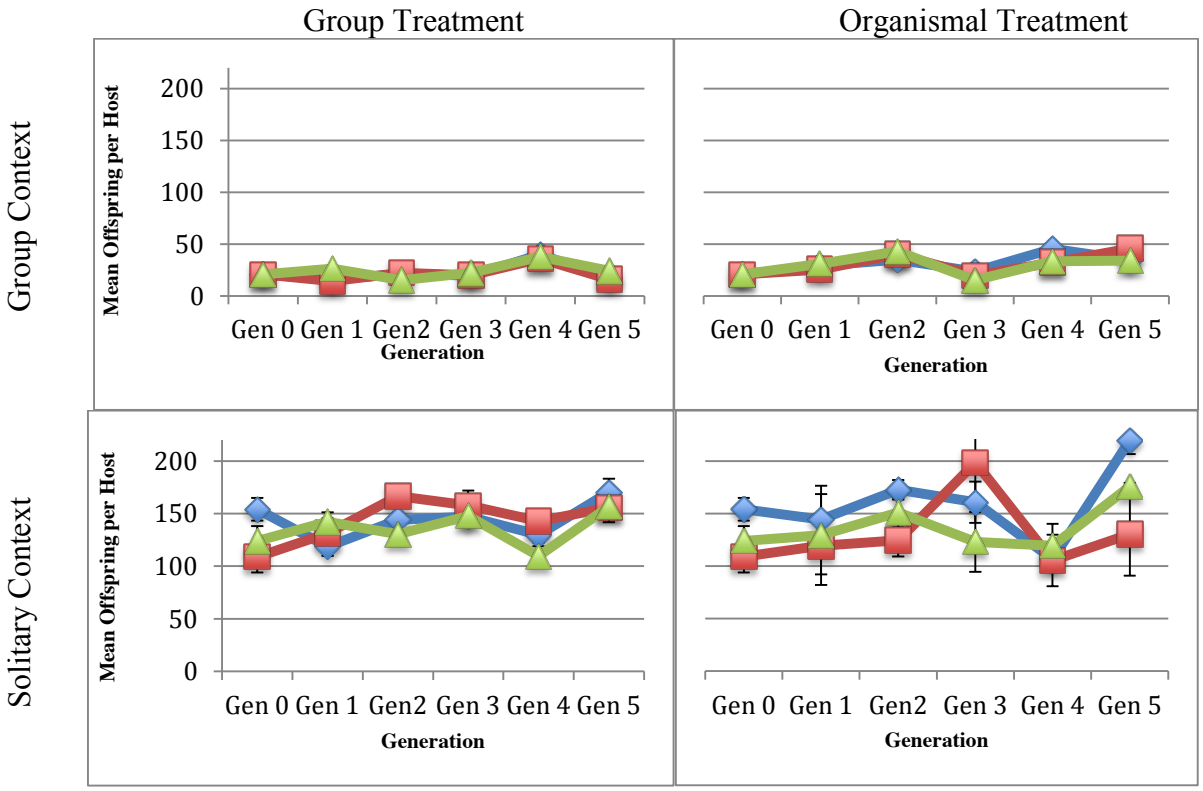


Figure 1.3. Population Size Results.
Population size represented by mean number of offspring per host for each replicate shown in each combination of treatment and context. Symbols/colors represent each replicate. Error bars show SEM; any values where error bars are not obvious fell within the symbol.

Sex ratios: In the group treatment, there was a statistically significant difference in the sex ratios of the collective context between Generation 1 and 5, with an increase in mean number of males per group (t=2.44, df=22, p=0.02). However, the solitary context of the group treatment did not show a significant change in sex ratio (t=0.63, df=104, p=0.53), nor did the solitary context of the organismal treatment (t=0.29, df=20, p=0.77), or the collective context of the organismal treatment (t=0.49, df=4, p=0.65) (Figure 1.4).

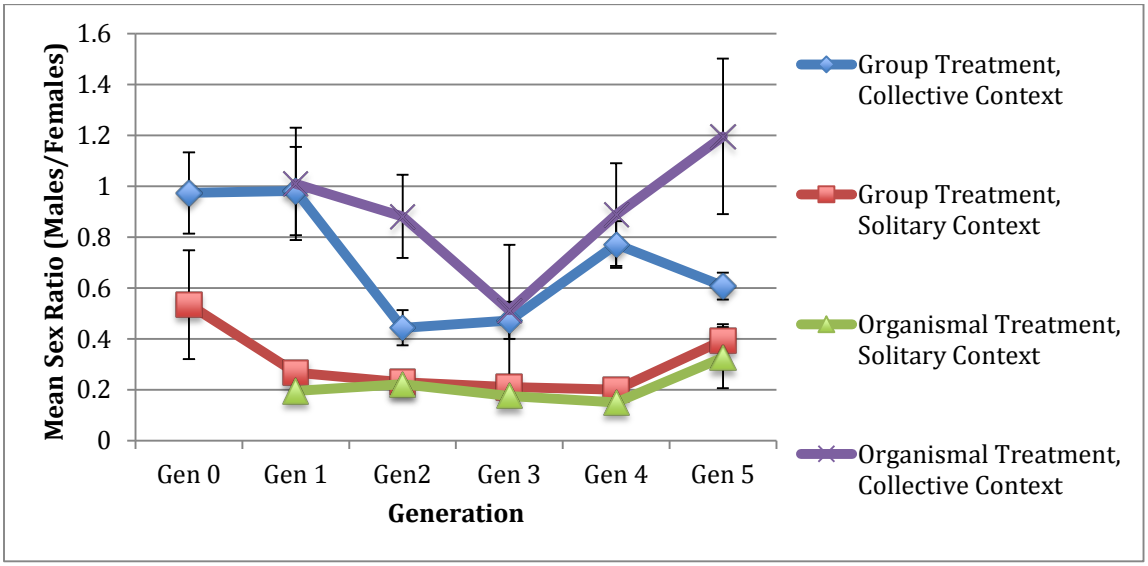


Figure 1.4. Sex Ratio Results.
Mean sex ratios for collective and solitary contexts of the group and organismal treatments. Error bars show SEM; any values for which error bars are not obvious fell within the symbol.

Discussion

Population size: Under the experimental conditions described above, a collective of five female wasps housed together as a group did not respond to selection for increased population size, as standardized by number of adults per host. Two hypotheses could explain this outcome.

Hypothesis 1) The experimental design did not impose sufficient strength or duration of selection to elicit an observable increase in population size.

The populations in this experiment competed via differential reproduction/extinction of groups: the largest vial spawned all of the populations in the next generation, while the smaller four vials did not produce new groups at all. This experiment used propagule pools in the creation of subsequent generations rather than migrant pools, a distinction that served to clarify the concept of reproduction at the group-level (Wade 1978). A propagule pool is created by randomly collecting a set number of adults from a group that is the target of selection and generating a new group from each propagule, while a migrant pool is created by first merging the emigrating adults from all groups and then collecting from this merged pool to create new groups. Propagule pools are particularly effective at selecting upon non-additive genetic effects or epistatic effects at the group-level as a result of population structure – such as frequencies of interacting phenotypes – because they tend to maintain the population structure of each selected group. In contrast, migrant pools reshuffle any population structure each generation by first merging all individuals (Wade 1978). Yet, despite these conditions that were favorable for selection at the group level, there was no response in population size.

In further consideration of hypothesis 1, the results observed in this experiment are reminiscent of at least two other studies in which no response was observed.

Baer *et al*. (2000), studying *Heterandria formosa*, a live-bearing fish, reported no response after six rounds of selection and concluded:

> this result suggests that extinction, either random or due to selection on population-level traits, may be unlikely to produce an evolutionary response in the demographic properties of populations unless the populations involved are quite small for a number of generations and/or migration among surviving populations is low. (Baer *et al*. 2000:177)

However, Baer *et al*. intentionally used a migrant pool design to decrease inbreeding, which consequently removed any effect that population structure might have had on the heritability of the trait in question. Their design also imposed weaker selection than other group-level selection experiments by selecting on three of five groups each generation, leaving the possibility that their experimental design did not provide enough time for this treatment – under the established conditions of strength and duration of selection – to exhibit a response.

Bashey and Lively (2009), studied the endopathic nematode *Steinernema carpocapsae*, and reported a response to selection at the group level in one of three source populations, while the other two showed no response. Yet, the three populations were derived from a single stock and differed in only two ways: 1) when they were separated from the stock population, and 2) how many passages they had undergone since separation. This suggests that the amount of genetic variation present in the population at the initiation of each experiment had an important influence on whether the population responded to the selection pressure imposed (Bashey and Lively, 2009).

It is also worth noting that Lewontin (1970) suggested that group-level traits would evolve more slowly than traits at the organismal-level or below because of the greater generation time of group-level traits. However, with respect to the results of my experiment, despite the lack of a response in the number of offspring per host at the group-level, there was a response in the sex ratio of the group-level and a change in the

number of offspring per host of the solitary females in the group treatment. This suggests that the strength and duration of selection were sufficient to see a response in both the group and organismal treatments, which is further supported by the use of a propagule pool and the initiation of the experiment with the genetically diverse ASURx population.

Hypothesis 2) Five female *Nasonia* housed collectively do not act as a "group" with respect to selection on population size.

An analysis of hypothesis 2 depends upon the definition of "group." However, theoreticians, philosophers, and experimentalists have varied greatly in how they conceive of a "group," and few have provided explicit definitions for their differing conceptions (Table 1.5). For example, as shown by column 2 of Table 1.5, only three of the 16 papers reported provided an explicit definition for the concept of a "group." Note that for the experimental studies, merely describing the experimental design of the groups used did not count as providing an explicit definition of group, as the experimental design is merely an operationalization of a definition rather than an explicit definition in and of itself.

Hull (1980) recognized a related problem concerning definitions in the group selection controversy:

> From the beginning of the controversy over group selection, two quite different sorts of "groups" seem to have been intended: highly organized groups exhibiting group characteristics and organisms that happen to be located in close proximity to each other. (Hull 1980:312)

Of these two types, Hull thought that too many arguments focused on the later, when they should have focused on the former. However, some researchers have attempted to define such "group characteristics" based upon particular kinds of interactions among group

members. For instance, Sober and Wilson (1998:92) wrote: "a group is defined as a set of individuals that influence each other's fitness with respect to a certain trait but not the fitness of those outside the group." In other words, building upon Wilson's (1975) "trait group" model, Sober and Wilson (1998) conceive of a group as a collection of organisms that engage in particular kinds of fitness-relevant-interactions. Colum 3 of Table 1.5 summarizes instances where a definition depends upon interactions among organisms. For instance, some of the experiments likely involved interaction among organisms within groups, but these interactions were not quantified or discussed in relation to the study's questions.

Source	Definition implicit or explicit?	Interaction based?	Spatially isolated?	Referenced only as evolutionary unit
Wynne-Edwards 1962	No	Yes	Yes	No
Maynard Smith 1964	No	Yes	Yes	No
Lack 1966	No	No	Yes	No
Williams 1966	No	No	Yes	No
Lewontin 1970	No	No	Uncertain*	Yes
Wilson 1975	Yes	Yes	No	No
Maynard Smith 1976	No	Yes	Yes	No
Wade 1976	No	Yes	Yes	No
Goodnight 1985	No	No	Yes	No
Heisler and Damuth 1988	No	Yes/No+ (MLS1/MLS2)	Yes/No+ (MLS1/MLS2)	Yes
Sober and Wilson 1998	Yes	Yes	No	No
Baer et al 2000	No	No	Yes	No
Okasha 2006	No	No	No	Yes
Pigliucci and Kaplan 2006	No	No	Uncertain*	Yes
West et al 2006	Yes	Yes	Yes	No
Bashey and Lively 2009	No	No	Yes	No

Table 1.5: Definitions of "group" summarizing the variety of factors different sources have used to define a group. See discussion for more detail about each column.
*This source did not provide enough information to determine if spatial isolation is a necessary part of the definition provided
+Heisler and Damuth (1988) intentionally distinguish between two different kinds of multilevel selection, here referred to as MLS1 and MLS2

Other researchers have emphasized some degree of persistence and isolation as essential to a concept of group for group-level selection. For instance, Maynard Smith (1964) used groups in his well-known haystack model that persisted separately for multiple generations so that competing alleles would become fixed in different groups. Column 4 of Table 1.5 shows which sources also rely upon spatial isolation to define "group." For experimental studies, this included experimental designs with separated groups.

Finally, some researchers have forgone defining the group itself, and instead focused exclusively on the types of features necessary for a group to serve as an evolutionary unit in the same way that organisms or other levels of organization might. For instance, in a quote reminiscent of Lewontin (1970), Pigliucci and Kaplan (2006:67) wrote:

> when selection 'sees' features of a group in a way that makes the features of the individual organism that make up the group invisible to the physical processes involved, it makes sense to think in terms of the group being selected.

This is to say that Pigliucci and Kaplan view a group, for the study of group-level selection, as defined by traits that—in Brandon's (1982, 1990) language—"screen off" selection at the level of the group. Column 5 of Table 1.5 denotes sources that were not exclusively addressing group-level selection, and mention the level of the group as one of multiple possible levels at which selection might act. Perhaps because of this, none of them provide definitions of "group" and seem instead focused on describing the features necessary for any level of organization to serve as a level at which selection might act.

Previous studies of *Nasonia* provide evidence suggesting the potential for a group-level trait such as "interaction among individuals," that would fit with some of these definitions. For instance, Burton-Chellew *et al*. (2008) proposed that in *Nasonia* females, decisions regarding their broods' sex ratios are largely mediated by the number of eggs already in a host, rather than through the physical presence of other females. Studies also show that different strains of *N*. *vitripennis* have heritable differences in the degree to which females adjust the sex ratio of their broods (Parker and Orzack 1985; Orzack *et al*. 1991), and females also facultativly adjust the size of their brood based on the number of times a host has already been parasitized (Burton-Chellew *et al*. 2008).

In my experiment, the wasps in the solitary contexts – regardless of selection treatment – produced significantly more offspring than wasps in the collective contexts. This suggests that the wasps in collectives were interacting in fitness relevant ways, such as through interference or competition (e.g., as in groups of chickens, Muir 1996), or females laid fewer eggs in the presence of other females since females preferentially lay eggs on unparasatized hosts (Ivens *et al*. 2009). The latter is likely a mechanism for conserving eggs for a possible future encounter with a host when fewer females are present. Either of these behaviors could account for fitness relevant interactions, which would make the collectives of five wasps a "group" *sensu* Sober and Wilson (1998).

However, because the collectives only existed for a single generation, they did not provide the persistence required to fit Maynard Smith's concept of "group" for group-level selection. Finally, because the number of offspring per host changed in the solitary context of the group treatment when the group-level trait did not, some might argue that

the collectives in this experiment were not groups *sensu* Pigliucci and Kaplan (2006) because there was no obvious trait at the group-level that screened off selection.

Sex ratios: In this experiment, *N*. *vitripennis* females responded directly or indirectly to the presence of other females and altered the sex ratios of their broods, an outcome expected based on previous studies (Warren 1980; Parker and Orzack 1985; Orzack *et al*. 1991; Burton-Chellew *et al*. 2008; Grillenberger *et al*. 2008). This result reinforces the earlier conclusion that the experimental design was appropriate for eliciting selection on some traits. When selected for increased number of offspring per host in a solitary context, the number of offspring per host increased, supporting the conclusion that even in laboratory strains there remains heritable variation in fecundity (Parker and Orzack 1985; Orzack *et al*. 1991). The proportion of males in each group also increased significantly in the solitary contexts of both the organismal selection and group selection treatments.

In addition to the changes in sex ratio in response to selection on offspring per host, this experiment suggests that when determining the sex ratio for their eggs, female wasps respond to the presence or absence of other females in a way that is at least partially contingent upon previous experience. *N*. *vitripennis* strains vary in their ability to respond to the presence of other females (Parker and Orzack 1985; Orzack *et al*. 1991), and these results suggest that *N*. *vitripennis* females experience a priming effect when determining how many eggs to oviposit on a particular host. In other words, it appears that females lay broods with more males when initially exposed to other females or to the presence of already-parasitized hosts—in this case, the transition from collective context to solitary context—while females lay broods with fewer males when previously exposed

to a solitary context and unparasitized hosts, such as the transition from the solitary context to the collective context.

Though my experiment shows significant differences in the aggregate sex ratios of the same wasps when laying in solitary and group contexts, the experiment does not provide direct evidence for how individual females are changing the sex ratio of their broods, since this study does not track the contribution of individual wasps to the group sex ratio, which is important for understanding the behavior of individual wasps (Orzack *et al*. 1991).

This observation might also provide an additional explanation for the lack of a response to selection at the group level, as it suggests that the wasps have behavioral adaptations that actively discourage forming integrated groups. This makes sense from the perspective of the individual female: in solitary contexts the female has greater access to resources for her offspring and will lay more females as expected from sex ratio theory (Hamilton 1967). However, when in the presence of other females, or when parasitizing patches that have already been parasitized by other females, a female wasp will lay a brood with a male-biased sex ratio, and will lay fewer eggs so as to conserve her investment for the possibility of finding another unparasitized patch. This would explain higher broods per host in both solitary contexts, the possible priming effect seen on the brood sex ratios, and would place selection pressure on mated females to avoid the presence of other females whenever possible.

Conclusion

I conducted a selection experiment with the parasitoid wasp *Nasonia* to test if five females housed in a single vial would respond to selection for increased population size.

Following similar designs in other studies of group-level selection (e.g., Wade 1976, 1977), the target of selection was the "group-level" collective of five wasps. I predicted that there would be a response to selection in the form of an increase in mean population size, but no such increase was observed in the group treatment.

I also predicted that the mean sex ratios would show an increase in number of males relative to females as part of the overall response to selection. In this respect, the results of the experiment did match with predictions, as the sex ratio of the group treatment did change. This result is consistent with previous research, and suggests that there was sufficient time and strength of selection to elicit a response under the conditions of this experiment. The results of this experiment also suggest that female *Nasonia* adjust the sex ratio of their broods in ways that partially depend upon conditions that they have experienced previously: females increase the relative proportion of males in future broods after they have encountered other females, or decrease the proportion of males in future broods after initially encountering no other females.

In discussing the results of this experiment, I alluded to the challenge of designing a multilevel selection experiment that is capable of comparing group and organismal selection pressures, particularly when using a parasitic system because of the additional challenge of standardizing available resources. Nonetheless, such experiments will be essential if we hope to continue testing conditions under which selection at levels other than the organism might operate.

However, the greater challenge in interpreting the results of this experiment became apparent as I considered the extent to which any interpretation was predicated upon how one conceives of a "group." That is, by some definitions of "group" this

experiment was an example of group-level selection that failed to result in a response to the target of selection, while by other definitions of "group" this experiment failed to include group-level selection at all. That such varied interpretations arise from different concepts of a group suggests the importance of understanding the history of conceptual change in the debates about group-level selection and for science in general. This conclusion also demonstrates the importance of experimentalists describing in detail the concept upon which their experimental designs and interpretations rest because, as Okasha wrote:

> Obviously, empirical data is crucial for resolving the levels-of-selection question, as for all scientific questions; but conceptual clarity is a prerequisite too. Unless we can agree on what it means for there to be selection at a given hierarchical level… then there is little prospect of empirical resolution, however much data we collect. (Okasha 2006:2)

Future experiments designed to select upon traits at the group-level, including those that elicit a response and those that do not, will aid our understanding of the conditions under which groups of organisms will or will not respond to selection at the group-level. Understanding these conditions for group-level selection will then aid in our broader understanding of how selection acts across multiple levels of biological organization.

CHAPTER 2

WHY ARE WE STILL ARGUING ABOUT GROUP SELECTION?

Abstract

During the 1960s, the long-standing idea that particular traits or behaviors could be explained by natural selection acting on traits that persisted "for the good of the group" prompted a series of debates about group-level selection and the plausibility of the idea that natural selection could act at or across multiple levels of biological organization. Even today this topic remains a contentious one, despite the fact that many researchers consider the debate as having already been settled. In 1983, D.S. Wilson reviewed the history of the group selection controversy, highlighting the ways that the debates had changed since the 1960s, and identified new evidence and existing conflicts related to the conceptual, empirical, and methodological aspects of these debates. By comparing the current literature on multilevel selection with the research that Wilson (1983) described, I summarize how the group selection controversy has and has not changed over the past 50 years. Major changes include the addition of new conceptual models and new empirical evidence, while the ways in which the current controversy remains largely unchanged are particularly related to methodological disputes. Finally, I argue that the history of "the" group selection controversy is best understood not as a single debate, as it is often presented, but as a series of distinct, though interrelated, debates within and across time periods. By understanding the history in this way, it becomes clear why some researchers have concluded that "the group selection debate" ended at various times, while other investigators continue arguing about the subject today.

Introduction

During the 1960s, the long-standing notion that particular characters or behaviors could be explained by natural selection acting on traits that persisted "for the good of the group" prompted a series of debates about group-level selection and the plausibility of the idea that natural selection could act at or across multiple levels of biological organization. For some researchers this topic remains contentious, yet other researchers consider the controversy as settled even while disagreeing about when and how this resolution occurred (Hamilton and Dimond 2012). Such disagreement about the current status of group selection thinking raises two important questions: Why are some researchers still arguing about the effectiveness of natural selection acting at one or more levels of biological organization? And why do researchers vary in their perception of whether and when a resolution to this controversy occurred**?**

To answer these questions, and simultaneously provide a review of the current state of multilevel selection, I use an approach similar to the "keyword" method used by Helmreich and Roosth (2010) and Keller and Lloyd (1992). In effect, this method follows the history of key ideas as a means of understanding the current controversy and theoretical framing. As Keller and Lloyd described:

> Attending to the multiple meanings of key terms provides a lens through which it is possible not only to understand better what is at issue in particular scientific debates but also to scrutinize the very structure of the arguments under debate. (Keller and Lloyd 1992:4)

To do so, I first identify a collection of core ideas at stake in the debates about group-level selection using David Sloan Wilson's (1983) review of the group-selection

controversy, and then compare these ideas with their current status in the literature on group and multilevel selection. My analysis illustrates how history can afford a way to understand the persistence of the controversy, and the failure to reach closure, by revealing how variation in the concepts, including the concept of a "group," influenced the interpretation of models, experimental design, and empirical data.

Wilson (1983) reviewed the then-current controversy intending to bring to light the complexity of the debates about group selection thinking and to summarize the most recent empirical and theoretical evidence in support of selection acting at the group-level. He did so with the aim of settling the ongoing disputes by providing a common historical and conceptual foundation for future work to build upon so that it would not be burdened by outdated criticisms. Yet, the controversy over natural selection operating at one or more levels of biological organization continues (e.g., Wild *et al*. 2009; Wade *et al*. 2010; Wild *et al*. 2010; Nowak *et al*. 2010; Abbot *et al*. 2011; Nowak *et al*. 2011).

I will show how the controversy over group-level selection is best understood through a lens of conceptual change. My thesis is that the history of "the" group selection controversy is best described not as a single debate, as it is often presented, but as a series of distinct, though often interrelated, debates driven by differences in what researchers have taken to be "the" group selection problem from the conceptual or theoretical standpoint. Further adding to the boundaries between these multiple debates, researchers have used varying forms of evidence in their explanations, often placing differing weights on the kinds of evidence used. By understanding the history in this way, it becomes clear why different researchers have concluded that the group selection debate

ended at different times, as well as why some researchers continue to argue about the effectiveness of natural selection at multiple levels of biological organization.

In exploring this thesis and the role of conceptual change throughout the sections of this chapter, I first provide a grounding in the controversies surrounding group selection in 1983 by reviewing Wilson's paper. I identify three broad categories – conceptual, empirical, and methodological – that organize how Wilson described the debates and how they changed during the history that he summarized. Then I use Wilson's (1983) review as a foundation to compare and analyze how the points Wilson addressed have or have not changed since the early 1980s, revealing the extent to which conceptual change has occurred by tracking the same organizational categories I identified in Wilson's analysis. Finally, I will provide my own historical analysis that demonstrates the usefulness of understanding "the group selection debate" as a series of interrelated debates often motivated by the questions and methods that researchers used that have influenced models, experimental design, and the interpretation of data.

Group Selection Before 1983

Wilson (1983) sought to both frame and settle "the group selection controversy," which by that point had been ongoing for more than 20 years. While natural selection was often seen as effective at multiple levels of biological organization before the 1960s (e.g., see Williams 1996, Collins 1986, Wilson and Wilson 2007), researchers and theoreticians during the 1960s raised arguments against many of the assumptions associated with group selection reasoning (e.g., Maynard Smith 1964, Williams 1966, Lack 1966; see Borrello 2010 for a review). By the late 1960s and early 1970s, biologists in general

accepted the argument that natural selection was not expected to be effective at the level of the group and higher. Wilson (1983) succinctly summarized the tenor of these times: "group selection rivaled Lamarkianism as the most thoroughly repudiated idea in evolutionary biology" (Wilson 1983:159). Later still, many would continue to think of group selection as a "solecism which would cause today's biology undergraduates to wince" (Dawkins, 2000: para. 3).

Despite this reputation, some continued working on the theoretical and empirical possibility of selection at the group level during the 1970s. For instance, Wade (1978) reviewed various group selection models from this time. One of Wilson's (1983) major arguments was that it was time to move past the previous conclusions that group selection lacked theoretical and empirical merit. To this end, Wilson's article had two sections. First, he described the history of the debates, highlighting the nature of previous models, especially their weaknesses, while arguing that the "new group selection" models that he and others developed during the 1970s did not suffer from the same failings. Second, Wilson provided a foundation for future research, explaining the then-current theoretical models and empirical evidence and suggesting new areas of research that group selection researchers could pursue.

In what follows, I review the historical portion of Wilson's paper, particularly the sections that provided the foundation for future research and the conflicts that Wilson argued were preventing researchers from moving beyond the controversy related to group level selection. I consider the topics that Wilson reviewed as falling into three areas: 1) the conceptual content of the debate, 2) the empirical evidence available, and 3) methodological disputes. I discuss each of these areas, including the topics within each

that Wilson (1983) described. Then I argue that even before the mid-1980s there was evidence of multiple interrelated debates about group selection rather than a single ongoing debate.

Conceptual content before 1983: new models and the concept of a group

New conceptual models: During the 1970s a new class of model describing how selection could act at the group level entered the group selection discussions: the intrademic group selection (IGS) model. Unlike prior group selection models that relied upon entire groups competing via differential survival, reproduction, or extinction, IGS models focused on the dynamics of organisms interacting within groups in ways that caused group-level properties to contribute to organismal-level fitness. In describing these models, Wilson (1983) aimed to counter two criticisms of group selection models:

1) Group selection models could not adequately represent a process by which natural selection at the level of the group would lead to a response or group adaptation.

2) The new—IGS—models might represent a natural process capable of leading to a response, but they shared nothing with the group selection models of the 1960s and were not representing selection at the group level.

To counter these claims, it was not enough for Wilson to simply describe the new IGS models. Instead, he first provided a review of older group selection models, both verbal and mathematical, to show that IGS models shared a theoretical coherence with the earlier group selection models. Then he described the application of these new models to show that IGS could effectively account for selection at the group level. Here I will follow Wilson's structure, first summarizing his review of what he called the "verbal

models" and "mathematical models" and then moving to a full description of the new IGS models.

Verbal models: Despite the fact that many of the models of group-level selection during the late 1970s and early 1980s focused on mathematical formulations of nature, Wilson reminded readers that these mathematical models were themselves "attempts to clarify a more nebulous, preexisting verbal idea" (Wilson 1983, 160). These earlier verbal models, Wilson argued, focused on a more "intuitive" sense of how adaptation generally worked, and these ideas were often influenced by societal attitudes rather than a deep understanding of the biology. In describing the verbal models of group selection, Wilson agreed with Williams (1966) regarding the role of social influences on how we might conceive of group-level selection and group-level adaptation:

> The conflicting perspectives of individual selfishness vs subservience to others are so pervasive among humans that it is hardly surprising that both found their way into evolutionary thinking and were originally stated as axioms. Thus individuals, groups of individuals, species, and entire ecosystems were all endowed with purposeful, organized behavior. (Wilson 1983:161)

In other words, because of the extensive cooperation that we see in human societies, it seems intuitive at some level to think about some traits as existing "for the good of the group" even in non-human species. This reasoning, by analogy, led to explanations of many apparently group-level traits as group-level adaptations.

But opponents of this "for the good of the group" view identified a key flaw in this line of reasoning. Wilson again pointed to Williams's (1966) description of the problem with group level thinking when applied to evolutionary biology:

> If natural selection favors individuals that leave the most offspring, then individuals that benefit themselves at the expense of others should be very

> fit indeed. Individuals that benefit others, presumably at some expense to themselves, would be selected against as surely as if they had bad eyes or faulty teeth. (Wilson 1983, 162)

This dilemma, the problem of how altruistic traits and seemingly maladaptive behaviors could persist in populations, became one major focus for the group selection discussions of the 1960s and 1970s. How can one explain apparently maladaptive behaviors while maintaining the framework of natural selection, which was understood to be an inherently selfish process? Explaining this dilemma, Wilson argued, was the goal of the verbal models of group selection. However, as I suggested in chapter 1 and discuss in more detail below, few of these verbal formulations provided detailed definitions for the concept of the "group" upon which selection was supposed to act. Perhaps partly in response to these nebulous formulations, the mathematical tradition took to explaining the same problems with more clarity and precision than the verbal models.

Mathematical models: In Wilson's (1983) account, three researchers' work, which I discuss below, became the foundation for different pathways to mathematically addressing the topic of group selection prior to 1972: Sewall Wright, Vero Copner Wynne-Edwards, and John Maynard Smith.

According to Wilson, "since the primary objection to the evolution of group-level adaptations was stated in mathematical terms, a mathematical answer was required; most people looked to the work of Sewall Wright" (1983:163). For Wilson, Wright's (1945) work provided a clear example of an early mathematical formulation of both the problem of explaining individually disadvantageous traits and the potential solution provided by between-group selection. The core of Wright's (1945) model was a single locus system with two alternative alleles. One of these alleles coded for a socially beneficial character

that was deleterious to an individual organism. Selection against organisms carrying this allele would either be (1-s) for heterozygotes carrying a single copy of the allele or (1-2s) for homozygotes carrying two copies of the allele, where "s" equals the selection coefficient. However, the benefit to the entire group would increase the fitness of all genotypes within the group by 1 plus the benefit multiplied by the number of beneficial alleles present. Thus, whenever the benefit to the group outweighed the cost to the organisms, the allele should encourage group growth, but because the allele is always selected against at the organismal-level, selection should tend to decrease the allele in the global population. Wright reasoned, however, that if these beneficial alleles were distributed across many populations in which the frequency of these alleles varied, then groups with the highest frequency of the beneficial allele would grow most quickly and create a larger number of new colonizing groups than other populations resulting in a form of between-group selection caused by differential growth and reproduction of the populations. Wilson (1983) pointed to Wright as concluding, "It is indeed difficult to see how socially advantageous but individually disadvantageous mutations can be fixed without some form of intergroup selection" (Wright 1945:417). Though, as I discuss at the end of this section, Wright himself did not seem to consider this process to be true "group selection."

Although V.C. Wynne-Edwards did not formulate a mathematical model of group selection, Wilson (1983) reasoned that so many mathematical models had been constructed based on Wynne-Edwards's (1962) verbal formulation that it was appropriate to describe Wynne-Edwards's work as a distinct source of mathematical models. Though Wilson (1983) seemed to criticize Wynne-Edwards (1962) for only considering

populations of animals rather than other kinds of groups, this is because Wynne-Edwards's work was concerned with answering what he called "the population problem:" the fact that most populations of animals have the capacity to exhaust their available resources, and yet do not. He proposed that populations regulate their numbers through a series of "homeostatic mechanisms" including epideictic displays and other behaviors that would provide a visual indicator to group members of how large the population was, as well as dominance behaviors such as territoriality that would limit access to potential mates and thus limit population growth. A core component of his argument was that these mechanisms persist because of differential extinction of the groups in which the mechanisms fail, resulting in a form of selection at the group level for the maintenance of these homeostatic mechanisms (see also: Wynne-Edwards 1963, 1985, 1993). As Wilson (1983) noted, a large number of the mathematical group selection models of the time bore "the stamp of Wynne-Edwards's ideas" (1983:166), and Wilson (1983) cited at least 10 such models. Most of these models consist of multiple groups and a fixed number of possible habitats that can be occupied by a single group. Within each group, organismal behavior is modeled as controlled by a single-locus, with one allele serving as a "selfish" variant that ultimately increases to fixation. At the group-level, each group faces a probability of extinction that is related to the frequency of the selfish allele. Much of the variation in these models derives from the nature of this relationship between a selfish allele and the probability of group extinction. If the group occupying a particular habitat becomes extinct, then the site becomes available for colonization by other groups. Thus, following Wynne-Edwards's (1962) argument, group selection in these models is based on differential extinction.

In response to Wynne-Edwards's (1962) book, John Maynard Smith (1964) distinguished the group selection that Wynne-Edwards advocated and W.D. Hamilton's (1964) then-recently published concept of inclusive fitness. In an interview many years later, Maynard Smith recalled that he met with others, including David Lack, after reading Wynne-Edwards (1962) and decided that Hamilton's idea should be renamed "kin selection" to clearly distinguish it from Wynne-Edwards's "group selection" (Maynard Smith 1997). Maynard Smith (1964) was the first to use this new wording in print when he presented a model to demonstrate why kin selection was a valid explanation for seemingly maladaptive behaviors while group selection was not. Maynard Smith's model—commonly called the haystack model—consisted of a population of mice living in haystacks in which a single Mendelian locus controls behavior. Mice with the dominant "A" allele are "aggressive" and continue to consume resources and reproduce even in populations with high densities, while mice homozygous for the "timid" "a" allele do not over-exploit resources and reproduce less as resources become scarce. Because of this, Maynard Smith argued, within any single group eventually the "A" allele will become fixed. However, at the population level, the "a" allele is favored because it leads to greater group persistence.

The model is initiated with a panmictic population containing an equal proportion of both alleles and then mated females disperse to found new colonies in separate haystacks. Each colony is then allowed to grow over a number of generations sufficient for the dominant "A" allele to reach fixation in any groups in which it is present. Finally, the offspring of all groups migrate back to the general panmictic population for another round of mating before starting again. Because of this, Maynard Smith (1964) argued,

any group that the dominant "A" allele entered would ultimately become fixed for the allele, and so selection for the organismal level selfishness would be stronger than the selection for the group-level benefits of the "a" allele except under rare conditions, such as very low levels of interbreeding between colonies. But as I will discuss later, Wynne-Edwards (1964) disagreed about the assumptions of this model and argued that when using biologically realistic assumptions the model results in group selection acting to preserve those groups that conserve resources. Notice also that Maynard Smith's model represents a migrant pool while Wynne-Edwards's model represents a propagule pool, a concept that Wade (1978) articulated to counter some of the confusion regarding how to represent group reproduction.

IGS models: Wilson (1983) argued that by comparing the similarities among these three distinct versions of mathematical group selection models, the continuity between them and the "new" IGS models developed during the 1970s becomes clear. Counting his own work, Wilson cited at least five different sources as having generated versions of intrademic group selection models: Price (1970, 1972), Matessi and Jayakar (1973, 1976), Charnov and Krebs (1975), Wilson (1975), and Cohen and Eshel (1976). However, it's not immediately clear that all of these were created to serve as intrademic models (e.g., Price 1970, 1972). Nonetheless, Wilson (1983) argued that while many of the specific details in these models varied, they were all built around the same core assumptions that allowed the models to work without falling to the criticisms leveled against the older formulations of group selection models.

While many the earlier group selection models were intended to represent persistent groups, the groups in intrademic selection models did not necessarily remain

cohesive over time. Instead, the IGS models start with a global population of organisms that then subdivides into local demes. While this subdivision does not alter the global allele frequency of the trait in question, it does create the potential that the allele frequencies within each group vary from the allele frequencies within the global population. Depending upon the details of the specific model, this formation of demes could be random or non-random, affecting the degree of variation between groups accordingly. After local demes are formed, selection then works within the demes until the organisms within each deme finally disperse back into the global population.

The ephemeral nature of these groups prompted some to argue that they were not really groups at all, but Wilson (1983) disagreed and argued that the IGS models shared a great deal of continuity with the more traditional group selection models. As Wilson correctly showed, the core of these intrademic models is simply a generalized version of Maynard Smith's haystack model, allowing for a relaxation of some assumptions and thus more variation than the original. Wilson (1983) also noted that the intrademic models are generally consistent with Wright's work in which the fitness of an organism partly depends upon the group context. Indeed, Wilson argued that the IGS models share common features with all of the prior group selection models that he described, including the fact that within local groups the group-beneficial allele would be selected against, but that between the local demes, groups with the highest frequency of the group-beneficial allele would be selected. For an excellent comparison between the traditional and intrademic group selection models, see Wade's (1978) review. But as Wilson concluded, "no one familiar with the history of the subject can doubt that IGS models fall squarely within the group selection tradition" (Wilson 1983:170).

The concept of a group: As mentioned above, Wilson (1983) suggested that the term "group" was being used in a variety of ways within debates about group-level selection, particularly since some did not consider the IGS models to be modeling true groups, and Table 5 from chapter 1 further illustrates this point. For instance, some models—both verbal and mathematical—defined a "group" as necessarily persisting over an extended period of time in order for group-level selection to function. Wilson (1983) cited Wynne-Edwards (1962) and Maynard Smith (1964) as defining their groups as persistent entities: Maynard Smith's multi-generational groups needed to persist long enough for the dominant allele to drive the alternative altruistic allele extinct, while Wynne-Edwards's groups required a degree of persistence sufficient to allow for differential extinction, so that ephemeral groups would likely be selected against.

However, Wilson (1983) argued, requirements of persistence need not necessarily be inherent to the definition of selection at any level, as illustrated by the fact that many mathematical models of natural selection in general do not even consider duration. Other definitions required a degree of separation to identify groups, but the IGS models work equally well with discrete groups and neighborhoods—spatially overlapping groups in which a single organism can be part of multiple groups—and IGS models also work with groups formed by behavioral segregation rather than spatial separation. Thus, many of the IGS models work without the assumptions that Maynard Smith's haystack model required, such as fixation of each allele within groups between bouts of dispersion. In other words, Wilson argued that overly strict definitions of "group" limited the feasibility and applicability of group-level selection models. I should note here, as I describe in

more detail below, that Wynne-Edwards (1964) also disagreed with the degree of persistence required by Maynard Smith's (1964) haystack model.

Yet, in addition to the degree of persistence that Wilson (1983) highlighted regarding differences in the definition of "group," even during the 1980s there were other differences in definitions. As I showed with Table 5 in chapter 1, there were also differences in whether groups were defined by a particular degree of spatial isolation, or whether fitness relevant interactions – such as those used in IGS models – were necessary. For example, Maynard Smith (1976) admitted that his (1964) description of group selection had failed to emphasize particular important points, such as group extinction. Nonetheless, he remained adamant that:

> For group selection, the division into groups which are partially isolated from one another is an essential feature. If group selection is to be responsible for the establishment of an "altruistic" gene, the groups must be small, or must from time to time be re-established by a few founders. This is because in a large group there is no way in which a new "altruistic" gene can be established. …
>
> What I should have said in my 1964 paper, but did not, is that the extinction of some groups and the "reproduction" of others are essential features of evolution by group selection. If groups are the units of selection, then they must have the properties of variation, multiplication, and heredity required if natural selection is to operate on them. (Maynard Smith 1976:279)

Notice that Maynard Smith continued to define groups based upon a measure of spatial isolation, as well as a degree of persistence such that it made sense to think of them as capable of going extinct. This was in contrast with Wilson's (1983) view of groups, which – like the other IGS models – were not spatially isolated at all, and were largely ephemeral except for key fitness relevant interactions.

As I discuss below, many of these conflicting definitions persist in current discussions of multilevel selection.

Empirical evidence before 1983: experimental studies of group selection

In addition to the various theoretical models supporting the idea that selection could occur at the group-level, Wilson (1983) also cited the growing body of experimental studies demonstrating evolutionary responses to group-level selection in laboratory settings (e.g., Wade 1976, 1977, 1979, 1982; Craig 1982; McCauley and Wade 1980, Wade and McCauley 1980). Wilson highlighted Wade's experiments with the flour beetle *Tribolium* that demonstrated responses to group-level selection, even when selection at the group-level was opposed by selection at the organismal-level. Wilson also noted that Wade's work provided evidence that some traits typically considered to be the result of organismal-level selection can alter as a response to group-level selection. For instance, regarding Wade's results:

> It is interesting to note that the major effect of such extreme group selection was to alter rates of cannibalism, which is the very sort of behavior that Williams dismissed as "not requiring" group selection! (Wilson 1983:183)

Wilson also highlighted the then-recent work in quantitative genetics that was coupled with these experiments, which showed that variation between groups can be high even when colonized by unrelated individuals:

> This empirical result suggests that the genetics of real populations may be more conducive to group selection than has been indicated by single-locus models. Although the underlying theory still needs to be developed fully, Wade and Craig have suggested that group selection can operate on the nonadditive component of genetic variation that is not subject to individual selection. (Wilson 1983:184)

Wilson (1983) offered these findings as evidence to counter the claims that group selection could only work in highly related groups, which was often accompanied by the argument that in such cases evolutionary change was the result of kin selection rather than true group selection, a point I return to later in this chapter.

Methodological concerns before 1983: local verses global fitness and the interchangeability of models

Local versus global fitness: Wilson (1983) noted that when calculating fitness of a single allele, IGS models allow the calculation in two ways: 1) within each group or, 2) across all groups as "global fitness". Wilson argued that calculating allelic fitness with this global measure yields the same result as calculating allelic fitness when one ignores the deme structure and only calculates the average allelic fitness in the total population. These two ways of calculating the global fitness—one including deme structure and the other not—Wilson argued, were equally valid for addressing whether a character associated with the allele in question would evolve or not. However, Wilson reasoned that when a researcher uses the simple calculation of allelic fitness that ignores the deme structure, the analysis will fail to capture the effect that the relative fitness of the allele within each local group will have on the way that the global population evolves. In other words, if one's goal is to understand the selective pressures on the trait in question, then understanding the group structure is important. This is why intrademic models calculate relative fitness within each deme rather than at the global level. Not taking this into account, Wilson (1983) argued, might lead one to mistakenly attribute the evolution of the character in question only to a single level of selection and then conclude that group

selection was not present in a case where it actually was. This would also explain why some argued that group selection was never necessary to begin with, if they only considered evolution of traits estimated at the level of global populations.

The interchangeability of models: Because of the variety of definitions for the term group, and the way that the group can be lost within global population analyses, Wilson (1983) argued that many of the population models used in evolutionary biology should be recognized as actually modeling the global population, as I described. Within the global populations of these models are smaller groups that fit the definitions used by IGS models. As others before him, (e.g., Wade 1978), Wilson (1983) argued that the sibling groups of traditional kin selection models could be considered a subset of IGS models, whereby the siblings of a kin model constitute a special kind of group in which all of the members happen to be related. This means, Wilson argued, that group selection models and models based on inclusive fitness are simply different methods of analyzing the same process:

> Hamilton's [1964] inclusive fitness theory is basically a different method of analyzing the same process. It correctly predicts the final outcome but does not distinguish clearly between the opposing forces of group and individual selection, as they are defined within the group selection tradition. (Wilson 1983:177)

This is to say that both kinds of models can predict the same outcome, but inclusive fitness models do not distinguish between the forces of group-level and organismal-level selection.

Of course, Wilson acknowledged that "the idea that kin selection is a special form of group selection grates harshly on many of those who attribute great importance to the former and none to the latter" (Wilson 1983:178). For instance, some argued that because

the intrademic group selection models could be translated to kin selection and inclusive fitness models, they were not really group selection models at all (e.g., Alexander and Borgia 1978; Maynard Smith 1964; Treisman 1983). Others seemed offended by the calculations used (e.g., Charlesworth 1981; Dawkins 1979; Alexander and Borgia 1978), either arguing that calculating relative fitness within groups was an error, or that the "differential productivity of groups caused by random variation must be classified as individual selection, while variation caused by reproduction within groups must be classified as kin selection" (Wilson 1983:179).

Wilson also argued that while it is fruitless to claim that reproduction within groups or degree of relatedness of group members could be used to distinguish kin and group selection, this could not justify one set of models as superior to the other:

> The fact that inclusive-fitness models and IGS models are merely different ways of analyzing the same population structure does not mean that one will ever supplant the other. Both have their separate insights and applications. It is imperative, however, always to have the appropriate set of definitions in mind and not to use one set to argue against the other tradition. (Wilson 1983:179)

Thus, Wilson argued that these approaches can and should be used together so as to address a wider range of questions and research topics.

Evidence of multiple debates before 1983

Though Wilson (1983) provided a single history of the group selection controversy, there are hints that it was more than just a linear story even during the 1960s and 1970s. For instance, Wilson (1983) and others (e.g., Wade 1978) intentionally distinguished between the verbal and the mathematical models of group selection, and

although Wilson (1983) argued that they are of the same tradition, he did not disagree that the IGS models are different in key respects than those that came before.

Indeed, even in the three sources of the mathematical tradition that Wilson (1983) identified there are suggestions of more than a "single history" for the group-selection controversy. For instance, Wade (1976, 1977) clearly grounded his experimental work within Wright's shifting balance theory, while others such as Wynne-Edwards (1962) and Lack (1966) based their work on discussions of population regulation.

It is also interesting to note that some researchers, including Wilson (1983), have counted Wright's (1945) work as group selection despite Wright (1980) having complained about his work in 1945 being called group selection. Wright (1980) distinguished between what he called "intergroup selection" and the more commonly discussed "group selection," arguing that group selection for the advantage of the group to be a "fragile" process. Indeed, as Wright noted:

> Maynard Smith, Williams and Dawkins have all discussed group selection for group advantage at length, and all have rejected it as of little or no evolutionary significance. They seem to have concluded, however, that this warrants the conclusion that natural selection is practically wholly genic, as implied by Fisher's fundamental theorem. None of them discussed group selection for organismic advantage to individuals, the dynamic factor in the shifting balance process[,] although this process[shifting balance], based on irreversible local peak-shifts[,] is not fragile at all, in contrast with the fairly obvious fragility of group selection for group advantage, which they considered worthy of extensive discussion before rejection. (Wright 1980:841)

In other words, Wright (1980) argued that his "intergroup selection" was different from group selection because the products of selection benefited organisms within groups rather than the groups themselves, and Wilson (1983) agreed. Yet others, including Williams (1966) and Wade (1978), counted Wright's shifting balance as a form of

traditional group selection. It is also worth noting here that Wynne-Edwards also referred to his version of group selection as "intergroup selection" (e.g., Wynne-Edwards 1963). Of course, later work in multilevel selection, as I discuss below, used IGS models in a way that seems to fit with Wright's distinction, where the model focused on group-level selection benefiting organisms within each group rather than benefiting the group itself.

Wilson (1983) also drew attention to the differences between the various forms of averaging strategies and arguments about whether group selection or kin selection models provide better explanations of the evolutionary processes. In particular, he highlighted the fact that "group selection is a fascinating example of how scientific questions arise from unscientific attitudes and of how their development is often haphazard and unsystematic" (Wilson 1983:159). But by understanding the history as a series of interrelated debates rather than a single continuous argument, some of the apparent haphazardness is less confusing. After all, if one views the history as continuous when there are actually multiple debates occurring, particularly with some of the arguments during the same time period actually stemming from different traditions, then it seems rather natural to expect that the history would appear haphazard. In the following section I will describe the ways that the controversy surrounding group-level selection—particularly the conceptual, empirical, and methodological aspects—has and has not changed and, in doing so, the presence of multiple distinct debates will become more apparent.

Current Group Selection Debates and Multilevel Selection

Thirty years after the publication of Wilson's (1983) review, it is striking to note some of the ways that the topic, and associated controversy, has changed, including the

transition to the study of multilevel selection. It is equally important, however, to note that some of the aspects Wilson (1983) argued should change are still little different than they were thirty years ago, particularly those related to methodical disputes. In the sections that follow I discuss first the changes related to the conceptual content, then the changes with respect to the empirical evidence. Finally, I discuss the similarities with respect to the empirical evidence, followed by the similarities with respect to method.

Changes related to conceptual content: multilevel selection, major evolutionary transitions, and the concept of a group revisited

Multilevel selection: Of the various changes that have occurred in discussions about group-level selection, perhaps the most significant is the transition from framing the debate as group selection versus organismal selection to arguments about whether and how natural selection might work across a variety of levels of biological organization simultaneously. These ideas, now referred to as multilevel selection, become common during the middle and late 1980s (e.g., Arnold and Fristrup 1982, Heisler and Damuth 1987, Damuth and Heisler 1988) although similar views prevailed before the name "multilevel selection" was coined in 1988.

Lewontin (1970), for example, described three requirements for any entity—organism, group, species, etc.—to be capable of evolution as a result of natural selection: 1) variation across the population for a particular trait, 2) heritability for the trait between generations, and 3) fitness consequences of the trait. Organisms often exhibit these characteristics, Lewontin (1970) argued, but any other level of organization with these three features should also be considered a potential "unit of selection." This included

levels of organization ranging from the cell to the organism to the group and the species. However, unlike later discussions of multilevel selection, Lewontin (1970) argued that those levels of organization above the organism would respond to selection more slowly than those of the organism or lower because their generation times would be much longer. Thus, Lewontin thought that the response to selection at the lower levels of organization would outpace any response to selection at higher levels. Because of this, he argued that levels of organization above the organism were likely to be poor units of selection in general, though he allowed for the possibility of exceptions, such as the level of the species. The later discussions of multilevel selection (e.g., Arnold and Fristrup 1982) would focus on understanding selection acting simultaneously across levels, including cases of selection at multiple levels working in the same direction, rather than only considering cases of conflicting selection and generally discounting the possibility of selection above the organism.

Damuth and Heisler (1988) first used the phrase "multilevel selection." Citing Arnold and Fristrup (1982) as the origin of the basic idea of multilevel selection, Damuth and Heisler (1988) explicitly connected the group selection debates with arguments about whether selection might work on entire species. They claimed that because both areas of research applied the principles of natural selection to levels of organization above the organism, "this leads them both to consider *multilevel* selection situations, where selection is occurring simultaneously at more than one level" (Damuth and Heisler 1988:407, original italics).

However, Damuth and Heisler (1988) acknowledged that there might be some challenges in unifying the study of multilevel selection because of differences in both

goals and available data. One challenge they singled out was the fact that researchers using the same terminology were sometimes referring to different processes, a point that Wilson (1983) also noted. Damuth and Heisler (1988) addressed this by offering a distinction between two different kinds of multilevel selection: "*multilevel selection [1]* and *multilevel selection [2]*" (Damuth and Heisler 1988:410, original italics and brackets). To better explain these, now often referred to as MLS1 and MLS2 respectively, consider their application to group-level selection.

The MLS1 study of group selection focuses on how organismal fitness is influenced by membership in a particular group. This means that fitness is counted at the level of the organism, but can be influenced by group-level properties of the group to which the organism belongs. The IGS models that Wilson (1983) described are examples of MLS1 models.

The MLS2 study of group selection focuses on the fitness of entire groups, often as measured by group reproduction. This means that groups are serving as individuals in the sense of Ghiselin (1974) and Hull (1976, 1978). As a consequence of this, for MLS2 the only traits of interest are those that belong to groups and the only form of reproduction that counts is when a group creates new groups. The group selection that Wynne-Edwards (1962, 1963) proposed corresponds to MLS2 models.

Damuth and Heisler (1988) recognized that their division of these two forms of multilevel selection was not the first, but they highlighted the fact that all of the previous divisions of these two versions of multilevel selection—be they from the group selection or species selection discussions—used the distinction to argue that one was correct while the other was somehow wrong. Indeed, Wright's (1980) distinction between "group

selection" and "intergroup selection" which I discussed above is one such example of this. Damuth and Heisler (1988) instead argued that both MLS1 and MLS2 are equally valid because both processes are present whenever multilevel selection is at work: "Once one has decided to analyze a given situation in terms of multilevel selection processes both approaches are legitimate within that context and a choice has to be made depending upon what questions are of interest" (1988:411). In other words, Damuth and Heisler (1988) argued that the particular questions of the research project should determine whether to use MLS1 or MLS2 because the focal unit of the analysis is different between the two, with MLS1 focusing on the organism within a larger collective and MLS2 focusing on the collective or group.

One strength of the multilevel selection discussions is the push for generality. The multilevel selection framework takes the core elements of natural selection and works to conceptualize them in a way that will apply across all levels of biological organization. In this sense, group selection is still discussed as a particular level of organization, although much of the work—both theoretical and empirical—is now applicable to multilevel selection in general. As I discuss later, however, there remains confusion in debates about group-level selection that would benefit from the MLS1/MLS2 distinction.

Major evolutionary transitions: Along with the distinction between MLS1 and MLS2, the concept of a major evolutionary transition helped to change the ways many researchers discussed multilevel selection during the 1990s. Maynard Smith and Szathmary (1995; Szathmary and Maynard Smith 1995) proposed that there have been eight major evolutionary transitions, each of which involved a change in the way that information is stored and transmitted. For example, the transition from entities using

RNA as the genetic material to the use of DNA as the genetic material marked a shift in the way that entities were transmitting information from one generation to the next. Other examples include the transition from unicellular organisms to multicellular organisms—which required the development of cellular differentiation such that only particular cell lines within the organism were involved in reproduction—and the analogous transition from solitary organisms to large colonies of organisms with division of labor and non-reproductive castes.

The major evolutionary transitions have become an important part of the multilevel selection discussions because many of these transitions involve collections of entities integrating into a higher-level entity upon which selection acts. Change in information transmission is seen in the fact that most of the lower-level entities are no longer directly involved in the reproduction of the higher-level entity. For instance, in the transition to multicellularity, most of the cells in a multicellular organism are not directly transmitting their genetic material to the next generation of the multicellular organism. This suggests that natural selection moved from acting upon the cells as separate entities to working on the collective cells as an individual. Understanding how such transitions occur fits well with MLS questions: when and how does selection on a group or collection of individuals result in the creation of a new individual that subsumes the members of the collective as components (Reeve and Keller, 1999; Okasha, 2006)?

The concept of a group revisited: While still a subject of discussion, the challenge regarding what should count as a group for the purpose of group-level selection has changed with the rise of multilevel selection. The subject now connects multiple topics, including discussions of individuality that arose in the context of species selection

(e.g., Ghiselin 1974: Hull 1976; see also, Vrba 1984; Gould 2002; Joblonski 2008), the units of selection as visualized by Lewontin (1970), and the more generalized replicator/interactor distinction (Dawkins 1982; Hull 1980). Because the new challenge is to define and identify the kinds of entities capable of evolving via natural selection—at whatever level of biological organization they may be—the particular issues that Wilson (1983) addressed are now part of much broader discussions regarding what biological entities are capable of evolution. This point is illustrated by the fifth column of Table 1.5 from chapter 1, showing definitions of "group" that barely mentioned the group at all, since they instead focused on describing general evolutionary units and mentioned that the group might serve as such a unit or level. For example, Okasha (2006) used the terms "particle" and "collective" to distinguish between nested entities:

> Intuitively, if selection is to operate at multiple levels, and lead to evolution, then entities at each level must satisfy the three requirements [of character variation, associated differences in fitness, and heritability]. Thus in a two-level scenario, the particles must vary with respect to a heritable character and differ in fitness as a result; and similarly for the collectives. If this is right, then the essence of multilevel selection is the simultaneous existence of character differences, associated differences in fitness, and heritability at more than one hierarchical level. (Okasha 2006:47)

Perhaps the most influential of these generalized concepts were developed by Dawkins (1976, 1982) and Hull (1980). Dawkins (1982), in presenting a general framework to discuss how selection acts upon a hierarchical system, framed the distinction as one between "replicators" and "vehicles." Replicators, Dawkins argued, are the units that actually reproduce and are passed from one generation to the next while vehicles are higher-level entities upon which selection directly acts. Following this distinction, Dawkins (1976) argued that genes are replicators and that organisms, and

perhaps even groups or species, are simply vehicles. Hull (1980, 2001) disagreed that there was a "natural division" of gene/organism/species as used by Dawkins and instead preferred the wording of "replicator" and "interactor" so as to not presuppose a specific organization of the biological hierarchy:

> If the traditional organizational hierarchy is retained, then both replication and interaction wander from level to level. The oblivious solution to this state of affairs is to replace the traditional organization hierarchy with a hierarchy whose levels are delineated in terms of the evolutionary process itself. (Hull 2001:23)

In other words, Hull argued that entities at multiple levels, rather than just genes, could potentially serve as replicators and the same should also be true of interactors. Thus, by the definitions of both Dawkins and Hull, the concept of replicator and interactor (or vehicle) was intended to be generalizable, such that replicator represents the unit of heredity, while the vehicle or interactor represents the entities that interact with the environment, both biotic and abiotic, so as to allow for differential replication (Dawkins 1976, 1982; Hull 2001). As such, rather than asking what kind of groups are capable of responding to group-level selection, they argued that the question is about whether groups are replicators or vehicles (Dawkins 1982) or even both (Hull 2001; Brandon 1999).

Brandon (1982, 1990), proposed a similar distinction with the terms "unit of selection" and "level of selection." Brandon later wrote that he was unaware of the Dawkins/Hull distinction, and thus his unit/level distinction follows the same logic: the unit of selection maps to the replicator while the level of selection maps to the interactor (Brandon 1999). In this sense, Brandon's definitions match Hull's most closely, since Brandon considered the possibility that some levels above the organism, such as

communities, might be both replicators and interactors. Note, however, that Brandon's usage of "unit of selection" is different than Lewontin's (1970), as what Lewontin described as a "unit of selection" Brandon called a "level of selection" (Brandon 1999). Adding to this potential confusion, other researchers prefer to use the terms "unit of selection" and "level of selection" interchangeably (e.g., Okasha 2006).

With discussions in multilevel selection during the 1990s and after, particularly those related to major evolutionary transitions, new questions were added to the challenge of how to define and identify units and levels of selection. Brandon's distinction was intended help researchers determine the particular level at which selection is acting. But along with understanding evolution as it occurs across multiple levels simultaneously, many MLS researchers have added the goal of understanding how higher-level entities can be formed from the integration of lower level-entities. That is, to use Hull's terminology, asking how selection can lead to the creation of new interactors (e.g., Michod and Nedelcu 2003; Okasha 2006).

Finally, the challenge of defining "individuals" in the sense that Ghiselin (1974) and Hull (1976) proposed, and identifying which levels of biological organization might count as individuals, has become even more complicated with recent realizations that many entities traditionally considered paragons of individuality—such as humans and charismatic megafauna—are themselves linked with diverse bacterial, fungal, and viral communities. For example, one recent study of the wasp *Nasonia* suggested some of the species boundaries are maintained by differences in the composition of their endosymbiotic communities (Brucker and Bordenstein 2013). Studies such as this are demonstrating a growing need to understand the function of holobionts: the combinations

of organism and their endosymbionts. The growing study of holobionts in humans, plants, and corals has even led some to argue that holobionts are themselves types of organisms (Gordon 2013), while others argue that we should stop thinking about genomes and focus on hologenomes as units of selection (Rosenberg *et al*. 2007; Singh *et al*. 2013; Gilbert *et al*. 2013; McFall-Ngai *et al*. 2013). Notice that in many ways these are conceptual extensions of earlier discussions, such as whether multispecies assemblages are interactors, as Hull might call them, or units of selection as Brandon described. The line of reasoning that holobionts might be units of selection also echoes arguments that entire communities of species might be units of selection (e.g., Lewontin 1970; Whitham *et al*. 2003; Whitham *et al*. 2006; Collins 2003).

Changes related to empirical evidence: new empirical studies of multilevel selection

In 1983 only a handful of experimental studies had explicitly tested the feasibility of group-level selection. There are now many more laboratory and field studies, including some framed to address multilevel selection and some still framed just as the early group selection experiments.

Following Wade's experiments in the late 1970s with flour beetles (1976, 1977, 1979), some researchers collaborated on the same system (Slatkin and Wade 1978; Wade and McCauley 1980; McCauley and Wade 1980) while others pursued their own group selection experiments (Craig 1982). In the 1980s, Goodnight, one of Wade's students, published the first group selection experiment using a plant system, demonstrating a response to selection on total leaf cover for small groups of *Arabidopsis* (Goodnight

1985; see also Goodnight *et al*. 1992). The 1990s saw the publication of an ongoing selection experiment in chickens, demonstrating a response to selection on the number of eggs laid in multi-hen cages (Muir 1996; Craig and Muir 1996). In the late 1990s, Goodnight and Stevens (1997) published a review of empirical group selection studies, including both laboratory experiments and quantitative fieldwork, nearly all of which showed a response to selection, followed by experiments in fish that showed no response to selection (Baer *et al*. 2000) and an experiment selecting on increased population size of parasitic nematodes that showed a response in some populations, but not others (Bashey and Lively 2009).

Since the early 1980s, studies have also been conducted on other levels of organization, such as two-species assemblages of *Tribolium* (Goodnight 1990a, b), multispecies assemblages of bacterial soil communities (Swenson *et al*. 2000a, b), and the communities formed by tree species with their associated arthropod inhabitants (Whitham *et al*. 2003; Whitham *et al*. 2006).

Most multilevel selection experiments demonstrate responses to selection at levels other than that of the organism (Table 2.1). Some of these cases demonstrate the importance of tracking multiple levels of selection to estimate the final result of selection by showing that both group and organismal selection can contribute to a single response. Yet, even the experiments that showed no response are useful for understanding the conditions under which particular levels of organization are likely to respond to selection.

Wilson (1983) argued that group selection as a theory should move toward developing predictions for when group-level selection will be overwhelmed by organismal-level selection. With the rise of multilevel selection as a topic of study this

same argument still applies, as we should be working to understand the conditions under which selection will act simultaneously across multiple levels of organization, or the conditions under which selection at one level will drive the evolutionary response and create units or interactors as described above. For instance, a recent series of experiments with single-celled yeast exerted a selection pressure by centrifuging and then propagating from the heaviest subset of the supernatant (Ratclif *et al*. 2013). These experiments have resulted in an otherwise single-celled variety of yeast that creates small multi-cell assemblages that then reproduce by budding new multi-cell assemblages. Under these selection conditions, where size is most important, the multi-cell groups outcompete even the largest single-celled yeast in what was called a demonstration of the transition from single cell to multicellular organisms (Ratclif *et al*. 2013).

Similarities with respect to empirical evidence: experiments show response to group selection

I have already described some of the new experimental studies published since the 1980s (Table 2.1). While the number of studies and types of systems studied has increased, the overall conclusion that, in the laboratory, group-level selection can be responsible for an evolutionary response remains strongly supported. Indeed, this conclusion has been supported again and again, even in the face of competing organismal level selection (e.g., Wade 1976, 1977; Goodnight 1985; Muir 1996; Craig and Muir 1996; Bashey and Lively 2009), and using either MLS1 or MLS2. Thus, the conclusion that there is experimental evidence for multilevel selection has not changed.

Table 2.1 – Experimental studies of group and multilevel selection published before 2000

Author(s), Year	Study Species	Common Name	Observed Response?
Wade 1976/1977	*Tribolium castaneum*	Red Flour Beetle	Yes
Wade 1979	*Tribolium castaneum*	Red Flour Beetle	Yes
McCauley and Wade 1980; Wade and McCauley 1980	*Tribolium castaneum*	Red Flour Beetle	Yes
Wade 1980	*Tribolium castaneum*	Red Flour Beetle	Yes
Craig 1982	*Tribolium castaneum*	Red Flour Beetle	Yes
Wade 1982	*Tribolium castaneum*	Red Flour Beetle	Yes
Wool 1982	*Tribolium castaneum*	Red Flour Beetle	No
Goodnight 1985	*Arabidopsis thaliana*	Thale cress (plant)	Yes
Breden and Wade 1989	*Plagiodera versicolor*	Willow Leaf Beetle	Yes
Garcia and Toro 1990	*Tribolium castaneum*	Red Flour Beetle	Yes
Goodnight 1990a, b	*T. castaneum and T. confusum*	Red Flour Beetles	Yes
Wade and Goodnight 1991	*Tribolium castaneum*	Red Flour Beetle	Yes
Muir 1996; Craig and Muir 1996	*Gallus gallus*	Domesticated Chicken	Yes
Swensen *et al*. 2000a	Unidentified	Microbial soil communities	Yes
Swensen *et al*. 2000b	Unidentified	Microbial soil communities	Yes
Baer *et al*. 2000	*Heterandria formosa*	Dwarf Livebearer (fish)	No

Similarities with respect to methodology: arguments against group selection, arguments for plurality, and confusion about the history

Arguments against group selection: Just as in 1983, current criticisms of group and multilevel selection come in a variety of forms. Note here that I mention both group and multilevel selection because, while some researchers level their arguments against the current conception of multilevel selection, many continue to argue against group selection without apparent knowledge of, or at least reference to, multilevel selection at all. Some researchers base their arguments upon current empirical or theoretical evidence, though many rely on arguments that are now outdated.

Eldakar and Wilson (2011) provided eight criticisms of group selection that are still used despite no longer being supported by current data or theory. Of the eight criticisms, six were criticisms that Wilson (1983) addressed either explicitly or implicitly (Table 2.2). For instance, criticism 1 is exactly the point that Wilson (1983) addressed as a problem of calculating the local versus global fitness of an allele, while criticism 3 is merely a variant of criticism 1. Indeed, not only was this point addressed in 1983, Wilson has continued to argue against this style of calculating fitness, which he and Sober have named "the averaging fallacy" (e.g., Sober and Wilson 1998; see also Okasha 2004a, 2006).

Table 2.2. Group Selection Criticisms Addressed by Eldakar and Wilson (2011)

Number	Criticism described by Eldakar and Wilson	Addressed by Wilson (1983):
Criticism No. 1	The fact that a trait evolves in the total population is an argument against group selection	X
Criticism No. 2	If a trait increases the absolute fitness of an individual, it does not require group selection to evolve	
Criticism No. 3	Conceptualizing the group as the social environment of the individual is an argument against group selection	X
Criticism No. 4	Frequency-dependent selection is an argument against group selection	
Criticism No. 5	The fact that a trait can be measured in individuals means that it evolved by individual-level selection	X
Criticism No. 6	Group selection is theoretically implausible	X
Criticism No. 7	There is little empirical support for group selection	X
Criticism No. 8	Group selection requires limited dispersal among groups	X

Not only have these now outdated criticisms plagued researchers attempting to work on the topic of group-level selection, the prevalence of outdated criticisms has fostered a sense of continued doubt about the validity of the theoretical framework in which the ideas of group-level selection are applied (Reeve and Keller 1999). As a consequence, some biologists simply avoid discussing group selection for fear of becoming entangled in the controversy, which is precisely one reason that Wilson (1983) wrote his review. For instance, compare Wilson's (1983) comment about researchers

avoiding connecting their work to group selection with a quote from Eldakar and Wilson (2011):

> So many people think of group selection as a bogeyman, however, that they are reluctant to accept any connection with their own favored ideas, and all efforts go toward finding differences. (Wilson 1983:180)
>
> Imagine submitting an article to a journal or a grant proposal to a funding agency, contemplating whether to explicitly frame your work in terms of group selection. Anticipating the comments that you are likely to receive, it is tempting to avoid using the term, even if group selection is clearly being invoked. Why not frame your work in terms of what evolves in the total population, the group as a social environment, or frequency-dependent selection instead? (Eldakar and Wilson 2011:1526)

Indeed, some researchers might actively avoid publically placing their work within a particular theoretical framework simply to avoid the possible controversy. For instance, during the symposium "In Light of Evolution V: Cooperation," Steve Frank (2011) gave an enlightening talk discussing pathology as a result of evolutionary conflict. In his talk, Frank described the tradeoff in microbes between rapid uptake of nutrients and efficient conversion to biomass. This rate versus yield relationship means that higher levels of relatedness tend to correlate with "cooperation" i.e., lower rates of consumption and thus higher total biomass yields. This idea could certainly fit within either a group selection or kin selection framework, but rather than identify with one or the other, Frank seemed to intentionally avoid the controversy: "call it group selection call it kin selection, it really doesn't matter."

That researchers would avoid using group selection phrasing, even when they admit that it could apply to their work, is perhaps unsurprising given the criticism that some have received when using such wording in the past. For instance, Wynne-Edwards was ridiculed for his group selection arguments and even into the 1990s he had difficulty

publishing some of his work on the subject (Racey 1997). Worse still, some contemporary views of group and multilevel selection are so polarized that both proponents and opponents claim that their view is the only proper and accepted approach. For example, compare this quote from Coyne's (2012) online writings with the following quote from Eldakar and Wilson (2011):

> Coyne (2012): The problem with all this is that the arguments for group selection are being made in books aimed at the general public, but the critical responses by evolutionary biologists are not only buried in technical papers, but involve arcane scientific arguments that sometimes use (horrors!) mathematics. So while group selection may flourish in the public mind, it's moribund to most evolutionary biologists who have followed the technical debates in the literature. (http://whyevolutionistrue.wordpress.com/2012/06/24/the-demise-of-group-selection/, accessed 9/5/2013)
>
> Eldakar and Wilson (2011:1526): Group selection is often portrayed as a subject that remains controversial after many decades. It would be more accurate to say that group selection remains confusing to many people after many decades. To anyone with a basic understanding of multilevel selection theory, the core question of whether a trait can evolve on the strength of between-group selection, even when selectively dis-advantageous within groups, was definitively answered long ago.

When biologists on both sides make claims that their argument is the only one supported by "most evolutionary biologists" or "anyone with a basic understanding of multilevel selection theory," then which side are those unfamiliar with the topic supposed to support? Given the factual nature of these claims, it is perhaps surprising that they should be so contradictory, as the truth is likely somewhere in between. Coyne and others are clearly wrong when claiming that no evolutionary biologists ever consider group or multilevel selection, as evidenced by the experimental studies mentioned above, while Wilson and others are so frequently in the habit of having to defend their views that they are perhaps overzealous in their responses.

Arguments for plurality: Wilson (1983) argued that inclusive fitness models and the group selection models were simply alternative ways of studying the same evolutionary processes. A key difference he added was that inclusive fitness "correctly predicts the final outcome but does not distinguish clearly between the opposing forces of group and individual selection, as they are defined within the group selection traditions" (Wilson 1983:177). But, because they were simply different approaches for analyzing the same process, Wilson argued that they were equally valid and both should be used. Wilson (1983) also argued that the study of evolution would benefit from a plurality of useful approaches to the problems that evolutionary biologists seek to explain.

Thirty years later, Wilson and others are still making these same arguments for a plurality of models and methods for studying selection across multiple levels of organization rather than supporting arguments that one should be used to the exclusion of all others. For instance, Sober and Wilson (1998) argued that as long as different alternatives lead to different useful insights, they deserve to coexist. Sober and Wilson (1998) also argued that researchers and theoreticians must make a distinction between process and perspective, by realizing that there are multiple perspectives, but that these various perspectives do not necessarily represent different processes. This means that the fact one perspective is able to explain the insights of the other is not necessarily a legitimate argument against either perspective:

> When one theory achieves an insight by virtue of its perspective, the same insight can usually be explained in retrospect by the other theories. As long as the relationships among the theories are clearly understood, this kind of pluralism is a healthy part of science. (Sober and Wilson 1998:98)

This kind of pluralism could play an important part in the development of evolutionary theory. Once this has happened, Sober and Wilson argued, a unified evolutionary theory for social behavior will emerge:

> We believe that a legitimate pluralism is possible and that it will lead to a unified evolutionary theory of social behavior. The theories that have been celebrated as alternatives to group selection are nothing of the sort. They are different ways of viewing evolution in multigroup populations. (Sober and Wilson 1998:57)

Others have provided similar arguments in favor of a plurality of perspectives to the development of evolutionary theory. For instance, Dugatkin and Reeve (1994:130) argued:

> Certain ecological scenarios are best approached from the individual selection road, whereas others are best traversed from the trait-group-selection path. Behavioral ecologists can only profit by having the 'nuts and bolts' of each approach in their conceptual toolbox.

Reeve and Keller (1999:4) also noted how alternative perspectives can be interchangeable:

> The particularly frustrating aspect of these constantly renewed debates is that, even though they seem to be sparked by rival theories about how evolution works, in fact, they often involve only rival metaphors for the very same evolutionary logic and are thus empirically empty.

Okasha (2006) provided even more detail on the topic of plurality, highlighting that there are actually a series of issues upon which a researcher might be a pluralist with respect to multilevel selection. Among these is pluralism about representation of the evolutionary process, which all of the researchers just mentioned seem to describe, as well as pluralism about causality of the process, and pluralism about the hierarchical organization of the systems we study. This last point seems to capture Hull's (1980, 2001) sentiments on the topic of the levels of selection. Finally, Bijma and Wade (2008)

represent a practical example of studying how multilevel selection, kin selection, and indirect genetic effects can all influence the overall response to selection.

One of the popular arguments against including multilevel selection in a plurality of perspectives to understanding the process of selection is the claim that relying upon multilevel selection as an explanation is unnecessary because kin selection can provide a superior account for any case that group or multilevel selection can explain. An example of this argument can be found in the "Social Semantics" series of exchanges in the *Journal of Evolutionary Biology* between West, Griffin, and Gardner (2007, 2008), and Wilson (2008).

West *et al*. (2007), responding to what they considered general semantic confusions within the study of social behavior, argued that the presence of different meaning for the same terms, and different terms for the same meanings, have hindered progress in evolutionary biology. They argued that such miscommunication "can obscure what is biologically important, and what is not" (West *et al*. 2007:415), including group selection which has "generated a huge amount of semantic confusion" (West *et al*. 2007:423):

> Although this debate was solved decisively during the 1960s to 1980s, by evolutionary biologists, it seems to reoccur and lead to confusion as new fields embrace the relevant aspects of social evolution theory.

In particular, West *et al*. highlighted the differences between what they called the "old" group selection and "new" group selection—essentially equating the verbal and mathematical models Wilson 1983 described as "old group selection" and the IGS models as "new" group selection—and argued that the potential validity of the "new," IGS, models might lead some to falsely apply the "old" models. Notice that they hint at a

variety of definitions for the concept of a group, but focus on differences in what is meant by "group selection" rather than the underlying issue of what counts as a group at all. I should also add here that West *et al*. (2007) only address group-level selection and make no mention of multilevel selection or Damuth and Heisler's (1988) work, despite the fact that this was exactly the type of semantic confusion that prompted Damuth and Heisler to distinguish between MLS1 and MLS2.

Wilson (2008:368), in an effort "to maximize the benefits of pluralism while minimizing the semantic costs" responded with a critique of West *et al*.'s (2007) article, particularly disagreeing with their characterization of group selection. While Wilson agreed with the goal of clarifying semantic confusion, he argued that West *et al*. failed to present a "genuine pluralism":

> Genuine pluralism requires crediting a given framework for achieving insights that were not forthcoming from other frameworks, even if they can be accounted for in retrospect. (Wilson 2008:371)

Wilson (2008) also argued that West *et al*. (2007) made three major errors in their analysis by, 1) presenting the "new" group selection—IGS—as having no historical connection with the "old" group selection, 2) presenting the rejection of group selection in the 1960s as proper and justified, and 3) arguing that the "new" group selection provides no insights that are not already provided by inclusive fitness theory. These points should now seem familiar, as they were largely the same points that Wilson (1983) raised 25 years earlier.

West *et al*. (2008) replied with a defense their original claims, as well as a challenge to Wilson's (2008) critiques:

> First, as we emphasized in our original paper, we do not claim that group selection is incorrect. Our point was that although it can be a potentially useful tool, it frequently leads to confusion and time wasting. Second, Wilson's paper is not a call for genuine pluralism; it is just a return to the old confusions about group selection, most of which were solved more than 20 years ago. (West *et al*. 2008:374)

This final point suggests that there are still confusions about the history of group and multilevel selection arguments, otherwise Wilson (2008) and West *et al*. (2007, 2008) would be unlikely to make conflicting claims about the history and status of the controversy surrounding group-level selection.

Confusion about the history: Wilson (1983) reviewed the group selection debates because he intended to illustrate key ways in which the debates had not changed. However, he was careful to add that he was not sure that what he presented was necessarily the best possible historical narrative, and he invited others to add to what he had started. Yet, by the late 2000s, a different kind of historical account had become the version often repeated in opening paragraphs about group selection. The following illustrates a more detailed version of what became a sort of "standard history" of the group selection debate:

> Darwin considered something like group selection to explain cooperation in humans and social insects, but didn't really invest in the idea; Kropotkin and Huxley had an involved debate that was thickly ideological; W. C. Allee and others at Chicago carried on for Kropotkin, sharing his preference for cooperation over competition; V. C. Wynne-Edwards started a cascade of events in 1962 that led first to powerful criticisms of group selection from G. C. Williams and John Maynard Smith and then finally to the end of the debate with the advent of Hamilton's (1964) inclusive fitness theory and the rise of sociobiology in the mid-1970s. (Hamilton and Dimond 2012:229)

Of course, there are many variants of this general account. For instance, here are three different versions, each of which seem to present the history as a single continuous story:

1) Dugatkin (2006) drew upon a history of altruism, which many consider to be the same history as the group selection controversy. In particular, he focused on seven individuals he took to be the most important: Charles Darwin, Thomas Huxley, Petr Kropotkin, Warder C. Allee, J.B.S. Haldane, and finally William D. Hamilton and George Price. Though Dugatkin mentioned other researchers, their roles were all abbreviated. Most importantly from a historical perspective, Dugatkin (2006) portrayed the people and events as a single story.

2) Leigh (2010) told the group selection story by opening with the standard account of Wynne-Edwards's (1962) book and its responses, but followed with a "prehistory" citing Wright (1929, 1932, 1945, and 1978) as invoking group selection. Leigh also cited Sturtevant (1938) for an argument noting that some traits can be favorable to the group but unfavorable to the individual and Lewontin's (1962) description of the t-allele system in mice, which Lewontin (1970) offered as a notable example of selection working at multiple levels. Most importantly, Leigh (2010) argued that there were two main reasons biologists invoked group selection: 1) to explain evolution of cooperation among individuals, as most use it, and 2) Wright's use of the spread of advantageous genotypes in complex epigenetic systems. In this sense, Leigh seems to follow the distinction that Wright (1980) himself made regarding his work and group selection, with the caveat that Leigh (2010) also considered it appropriate to describe Wright's characterization of how selection operated as "group selection."

3) Borello (2010) told a more complete story of the early disputes regarding the idea of group-level selection. He began with an argument that Darwin considered how natural selection could apply to levels other than the organism, such as Darwin's (1871) descriptions of how a group of men who cooperated would out-compete a group of men who did not. Borrello (2010) then followed the group selection history through Krotpotkin's (1902) view of cooperation as the driving force for evolution and Huxley's (1893) opposing view that competition is the main force of selection. Borello's (2010) main focus was V.C. Wynne-Edwards and his disagreements with David Lack, using the publication of *Animal Dispersion in Relation to Social Behavior* (Wynne-Edwards 1962) as the "structural spine" in an analysis of what motivated Wynne-Edwards' work, as well as demonstrating the wide range of reactions that the book provoked. For a more detailed review of Borello (2010), see Hamilton and Dimond (2012).

Notice that these examples, as do many others, fail to mention a few noteworthy events that occurred after the 1960s. In particular, some of the primary critics of group selection during the 1960s, including Williams and Hamilton, later softened their criticisms or even reversed their arguments entirely. For instance, in the reissued 1996 preface of Williams's (1966) book, he reflected:

> A few years after 1966, I was being given credit for showing that the adaptation concept was not usually applicable at the population or higher levels, and that Wynne-Edwards's thesis that group selection regularly leads to regulation of population density by individual restraints on reproduction was without merit. It also became fashionable to cite my work (sometimes, I suspect, by people who had not read it) as showing that effective selection above the individual level can be ruled out. My recollection, and my current interpretation of the text, especially of Chapter 4, indicate that this is a misreading. I concluded merely that group selection was not strong enough to produce what I termed *biotic adaptation*: any complex mechanism clearly designed to augment the

> success of a population or a more inclusive group. A biotic adaptation would be characterized by organisms' playing roles that would subordinate their individual interests for some higher value, as in the often proposed benefit to the species. (Williams 1996:xii)

Notice that the distinction Williams seems to be making here is essentially between MLS1 and MLS2, where MLS2 at the group-level would require identifying what he called biotic adaptations. But such a requirement does nothing to rule out the possibility of MLS1 group-level selection that would influence the fitness of organisms within groups.

Similarly, Hamilton described in his autobiography his initial dislike of the group selection idea until a phone call in which George Price asked if Hamilton had seen what Price's work, "a covariance formula the like of which I had never seen" (Hamilton 1996:173), did for group selection:

> I told him, of course, no, and may have added something like: 'So you actually believe in that do you?' Up to this contact with Price, and indeed for some time after, I had regarded group selection as so ill defined, so wooly in the uses made by its proponents, and so generally powerless against individual and genic levels, that the idea might as well be omitted from the toolkit of a working evolutionist.
>
> I am pleased to say that, amidst all else that I ought to have done and did not do, some months before he [Price] died I was on the phone telling him enthusiastically that through a 'group-level' extension of his formula I now had a far better understanding of group selection and was possessed of a far better tool for all forms of selection acting at one level or at many than I had ever had before. (Hamilton 1996:173)

There are, however, some authors who do not present the history of group and multilevel selection as a single story. For instance, Reeve and Keller (1999) while attempting to "make yet one more attempt to bury the issue that usually usurps discussions of the level of selection at the expense of the truly interesting issues" (1999:3), did not describe a single long history of group selection. Instead, they

characterized the controversy as a series of "cyclic debates" regarding the application of individuality to different levels in order to find the "true" unit of selection:

> This issue emerges in cyclic debates about (a) whether genes or individuals are best seen as the true unit of selection, and (b) whether groups of individuals *can* be units of selection. In our opinion, these questions have been satisfactorily answered repeatedly, only to reappear subsequently with naïve ferocity in new biological subdisciplines (e.g., the group-selection controversy is currently generating copious amounts of smoke within the human sciences; see, e.g., Wilson and Sober 1994 and responses; Sober and Wilson 1998). (Reeve and Keller 1999:3-4, original citations included)

Thus, in Reeve and Keller's (1999) account of the history, a series of debates begin and end before starting anew in a new field. But even within the human sciences, group selection had been discussed since the 1980s with the use of group-level selection in explaining cultural evolution (Boyd and Richerson 1985, 2005; Richerson and Boyd 2005; see also Henrich and Henrich 2007). Indeed, even the works that Reeve and Keller cite, Wilson and Sober (1994) and Sober and Wilson (1998), build upon insights provided by the human behavioral sciences (e.g., Boyd and Richerson 1985) as well as the then-current work in evolutionary biology. This demonstrates much more temporal overlap to the debates than Reeve and Keller (1999) described. This point also suggests, particularly when coupled with the wide diversity of the narratives provided above, that there are entire areas of MLS research not included in most of the historical account.

To explain these discrepancies, I argue that rather than a single continuous debate, or even multiple independent debates, the history of the controversy over group-level selection is better viewed as a series of interrelated topics and associated debates.

Multiple Histories

A first step in disentangling the issues surrounding the ongoing group and multilevel selection controversy is to appreciate that researchers have addressed the topic from a variety of contexts and have viewed different points as needing to be explained. For the differences in debates across time, the historical context of the time period in which the arguments were presented likely has a large influence on the kinds of questions that were asked. In contrast, for the differences in debates within particular times, the backgrounds that motivated the particular researchers might be better explanations for the differences in their questions and research.

For instance, consider the long-term disagreement between Wright (1929, 1932, 1945, 1978, 1980) and Fisher (1918, 1937, 1941, 1953, 1958) about whether it was better to characterize selection as acting on subdivided populations in which mating is more likely between individuals in the same deme or panmictic populations in which mating among all individuals is equally likely. As previously mentioned, many researchers have connected Wright's work to group and multilevel selection, but for Wright, the motivating issue was how gene frequencies might change as a result of selection or drift within and between subdivisions—or groups—that collectively composed a larger population. See Winther *et al*. (2013) for more of my work on the group selection controversy compared with the Fisher-Wright debate.

Contrast Wright's and Fisher's argument with those of Wynne-Edwards and Lack: Wynne-Edwards (1962) responding to Lack (1954), debated how best to explain the fact that some bird species were biologically capable of laying more eggs than were found in the average nest. Wynne-Edwards thought that selection at the group-level

provided an explanation, interpreting certain behaviors as evidence that birds were restraining themselves relative to the potential number of offspring they might produce. A similar example of restraint was the fact that some sexually mature individuals forego their own reproduction to assist parents or siblings in raising young. Lack (1966) responded to Wynne-Edwards's (1962) challenge by providing numerous arguments countering the examples Wynne-Edwards presented, arguing that the behaviors observed could be explained by the birds maximizing their reproduction relative to the resources required. Thus cases where younger birds forwent reproduction to help raise siblings could be viewed as individually beneficial with inclusive fitness models. It was only in Lack's (1966) appendix that he explicitly argued against Wynne-Edwards's (1962) group-level selection as theoretically ungrounded.

During the same time period, however, Maynard Smith (1964) responded to Wynne-Edwards (1962) with an article that renamed Hamilton's (1964) inclusive fitness approach kin selection and argued it was an alternative to group selection that remained consistent with individual selection. Williams (1966, 1996) argued that group selection, if even possible, was likely a weak force compared to selection at the organismal or genic levels. These two responses were aimed at the theoretical sufficiency of group selection rather than the empirical examples that Wynne-Edwards provided. Indeed, Williams was not even aware of Wynne-Edwards's book until he was nearly finished writing his own (1966) work (Williams 1996).

Finally, compare the above disputes with the exchange between West *et al*. (2007, 2008) and Wilson (2008), which is best understood as a conflict over the theoretical usefulness of group selection and the perceived costs of semantic confusion over the

associated words and concepts. Much of this disagreement also seems to be grounded upon factual disputes over the history of the group selection controversy.

These examples do not represent a single, continuous debate regarding group-level selection. Rather, they provide examples of debates on different topics largely occurring at different times, that all fall under the subject of group or multilevel selection. As such, they are not necessarily connected one to the other in the sense that one is an explicit response to another. Fisher and Wright argued about the best context for selection, a panmictic or subdivided population; Wynne-Edwards and Lack debated interpretations of data; Maynard Smith and Williams relied on mathematical and theoretical arguments against the plausibility of group selection occurring; while West *et al*. and Wilson disagreed about the usefulness of group selection or multilevel selection as a theory. Thus, rather than a single debate, as the history is often viewed, this brief account shows how the particular framing of questions by different researchers, as well as the evidence that they used, differentiated efforts to understand the level of biological organization at which natural selection acts.

Conclusions

Using "keyword" style methods, I have shown the ways central ideas at stake in the debates about group-level selection have and have not changed since the early 1980s. By viewing the history of group and multilevel selection as a series of interrelated debates separated by the ways questions were framed and the historical context within which research occurred, it becomes easier to understand why researchers disagree about whether "the" group selection debate was settled and why different researchers seem to

think it was settled at different times**.** Though the broad subject has remained largely the same—that is, exploring the possibility that natural selection can act at multiple levels of biological organization—many of the specific topics, as well as the associated questions and types of evidence used, have varied over time, which is illustrated by all that has changed in the discussions since the early 1960s.

In particular, I have shown how the conceptual and empirical aspects of these debates have changed: new models were proposed and debated, new empirical studies were conducted and interpreted, and new theoretical frameworks developed that provide for broader explanatory power. In contrast, conflicts that remain little changed are largely related to methodology, including differences of opinion about whether pluralism is appropriate, disputes about the best way to estimate allelic fitness, and even arguments about the basic usefulness of group or multilevel selection as a theory. Given these two points, it seems reasonable to wonder: How will these debates change in the future?

Wilson (1983) argued that group selection should be integrated with other major concepts in evolutionary biology, reflecting that,

> one of the most striking features of the "new" group selection [IGS] is its relation to other major concepts, such as inclusive fitness, game theory, and reciprocity. In the past these have been treated as rival theories, with every effort being devoted to accentuating their differences. Now it is apparent that they can be united within a single frame work [sic] and that far more is to be gained by emphasizing their similarities. (Wilson 1983:160)

For some, multilevel selection is that unification – or has the potential to be. However, as long as we continue to argue about the same details that have been debated since the 1960s, even the potential for a unification of theory will continue to be denied. As Reeve and Keller (1999) argued, it is time to move to more interesting questions. Viewing the

history of these conversations as a series of interrelated topics allows for the understanding that within these discussions new and interesting topics of debates can arise and then be settled over time. This is the style of debate that seems useful for the field, as it allows for building upon the arguments of the past while also supporting movement into new topics of discussion and debate. For instance, for what questions and under which biological conditions is it more productive to start with a multilevel selection model rather than an inclusive fitness model? Under what conditions might we expect that group-level selection will be a driver for evolution, and for which organizational levels, and under what conditions might we reasonably ignore it?

If we move toward these and other new questions, then the answer to the question posed in this chapter's title—Why are we still arguing about group selection? —becomes: "Because we are working to make the theory of multilevel selection even stronger."

CHAPTER 3

DIFFERING WAYS OF STUDYING MULTILEVEL SELECTION AND THEIR IMPLICATIONS FOR EXPERIMENTAL DATA AND THEORETICAL ARGUMENTS

Abstract

In this chapter I provide a framework for understanding the ongoing disputes about group selection, and multilevel selection more generally, by arguing that there are different "approaches" to the study of multilevel selection distinguished by where the research falls upon two axes: 1) whether the researcher is primarily interested in a single level of organization or multiple levels, and 2) the extent to which the work starts from a theoretical or empirical perspective. I then describe four such "approaches" including the questions asked and the methods used. I then discuss how these different approaches might lead to conflicting interpretations of the same empirical data by using the *Nasonia* experiment from chapter 1 as an example, discussing conflicts of interpretation that might arise between approaches and within approaches. I also discuss how the approaches described here can account for some conflicts of both methodology and epistemology. By acknowledging these different approaches, it becomes clear why some of the controversy surrounding group selection continues, and in doing so provides avenues for avoiding the same conflicts in future research.

Introduction

In chapter 2, I argued that the history of the group selection controversy—including the transition to discussions about multilevel selection—is best understood

through the lens of conceptual change as representing a series of interrelated debates separated by the topics and questions, as well as the favored methods and types of evidence, of researchers within and across particular times. In this chapter I ask whether this view of the history, and the lens of conceptual change, also provide insight into the contemporary controversy about group and multilevel selection.

Sober and Wilson (1994), in providing an overview of the philosophy of biology literature on multilevel selection, argued that an important part of the debate is properly framing the problem that needs to be explained before an appropriate explanation can be found. Following this reasoning, I propose that it is possible to separate studies of group-level selection using the intersection of two continua that represent different ways that researchers have attempted to frame the questions that are central to their research (Figure 3.1). I also argue that because of the differences in the kind of questions asked and general methods used—that is, theoretical or empirical—these alternative ways of framing the research questions may lead to differences in the interpretation of observational and experimental data.

For ease of reference I will hereafter refer to these alternatives as "approaches" to the study of multilevel selection. By approach I mean a particular combination of topic and questions along with the methods used to answer the questions. This use of the word "approach" is similar in connotation to the way that other researchers have used the word, such as Wilson (e.g., 1983) when he referred to group-selection and kin selection as "alternative approaches" to studying the same process, Goodnight and Stevens (1997) in describing the "adaptationist" and "genetical" approaches to multilevel selection, and Damuth and Heisler (1988) in describing MLS1 and MLS2 as different approaches to

studying selection in hierarchical systems. In other words, each of these uses of the word "approach" denoted—to varying extents—differences in the way that researchers framed the particular questions that they followed, as well as the concepts and methods that they used to address those questions.

In this chapter, I first describe the two intersecting continua of "single level or multilevel" and "theoretical or empirical." I then describe the resulting approaches to the study of multilevel selection, using the four most extreme as examples. Next, I use the results of the *Nasonia* experiment reported in chapter 1 to demonstrate how different approaches can lead to different interpretations of the same empirical data. Finally, I describe some of the conflicts that can arise between and within these approaches, including epistemological and methodological, which help to explain some of the ongoing controversy described in chapter 2.

Characterizing Research on Multilevel Selection

Though varied fields of research have included the study of multilevel selection, here I will focus primarily on the biology and philosophy of biology literature in three areas: theoretical evolutionary biology, philosophy of biology, and experimental evolutionary biology.

I propose that multilevel selection studies can be classified along two axes: 1) the degree to which the research focused on one level or multiple levels of biological organization and 2) the degree to which the research was primarily a theoretical or empirical analysis. Although both of these are broad continua, for ease of analysis here I divide them into discrete categories based on the ends of each continuum (Figure 3.1).

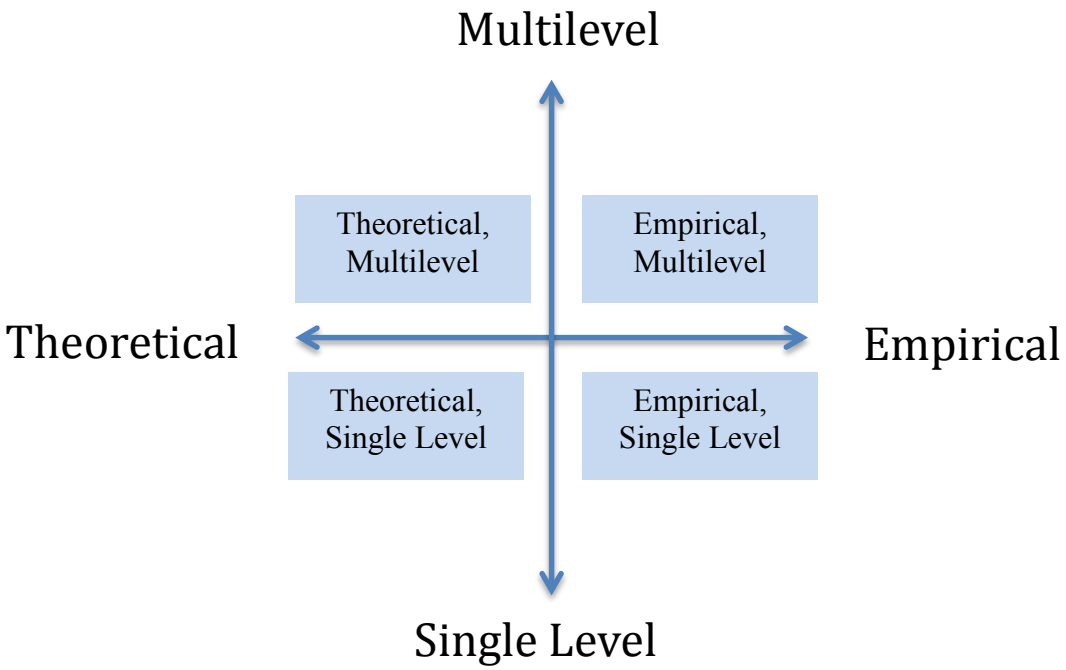


Figure 3.1. Representation of four approaches to the study of multilevel selection.

Single level or multilevel: Researchers vary in whether their research focuses mainly on selection acting at one level or selection acting at multiple levels simultaneously. I call the ends of this continuum "single level" and "multilevel." Researchers at one end are focused on understanding or identifying a level of selection most responsible for driving evolution – be it in general or for a particular trait. At the other extreme are researchers interested in understanding the dynamics of two or more levels of selection acting simultaneously that might influence evolution, including how to identify the multiple levels at which selection might act in general or the particular levels of selection contributing to the response of a given trait.

In some ways these extremes match the distinction between MLS1 and MLS2, including the fact that it is the kinds of questions that each address that drive the

distinction (Damuth and Heisler 1988:411): "Once one has decided to analyze a given situation in terms of multilevel selection processes both approaches are legitimate within that context and a choice has to be made depending upon what questions are of interest." With MLS1, researchers address how the fitness of entities at one level are influenced by the collectives of which they are a part; for instance, how organismal fitness might be influenced by traits at the group-level. With MLS2, researchers address how entities at one level respond to selection at the same level of organization; for instance, how group fitness should be calculated by the reproductive act of creating new groups.

However, the distinction between single level and multilevel that I use here is more than a renaming of MLS1 and MLS2. For instance, Wade's (1976) work could be considered MLS2 because the selection he exerted upon groups was done through differential reproduction and extinction of groups, yet his experiment included the ability to compare selection acting at multiple levels. This is why, in what follows, I describe his work as falling upon the multilevel end of the continuum. Equally, I can imagine the case of a researcher who is interested in whether selection upon an IGS modeled trait-group is a better explanation for a particular behavior than selection upon the organismal level. Though trait group models are often considered MLS1, this work would fall within the single level end of the continuum because the goal is in understanding the most important or dominant level of selection for a particular trait.

Theoretical or empirical: While, to some degree, all work on multilevel selection uses both data and theory, the extent to which particular researchers rely upon one or the other can be placed on a continuum between the ends of primarily theoretical research and primarily empirical research. Theoretical researchers start with abstract

concepts or *a priori* claims and then might investigate how well data fit with these conceptualizations. At the other end of the continuum, empirical researchers start by gathering observations or experimental results and then might move to connecting those results and observations to existing theory or constructing new theory. Thus, at its core, this distinction is about the degree to which a researcher starts with a general understanding of the natural world and then works to create theory for how to understand particular cases, versus starting with one or more particular cases of empirical studies and working from these cases to construct theory.

Goodnight (2005) and Goodnight and Stevens (1997), described a similar distinction between empirical and theoretical work by describing what they called the adaptationist approach, which is focused on explaining existing adaptations, and the genetical approach, which is focused on understanding ongoing changes either through experimental or observational studies. However, rather than two versions of study described by Goodnight and Stevens (1997), here I use these as two ends of a continuum, and so the distinction is not quite as sharp. Instead, the difference is in the degree to which data are used at the initiation of the project, or the degree to which the project starts from a theoretical perspective.

To illustrate the usefulness of dividing approaches to the study of multilevel selection, consider a brief example from the group selection controversy during the 1960s, focusing on four key publications that I described in chapter 2: Wynne-Edwards (1962), Maynard Smith (1964), Lack (1966), and Williams (1966). Because all four of these works were addressing whether group-level selection could be a driving force of evolution, they could all be categorized as single level (Figure 3.1). However, the

differences in the debates are easier to see when we note that Wynne-Edwards and Lack started with data before moving to generate theory, while Maynard Smith and Williams centered their arguments on theory rather than data. Thus, in key respects, Wynne-Edwards and Lack shared one approach, while Maynard-Smith and Williams shared another. As I also argued in chapter 2, Wynne-Edwards and Lack also shared particular questions, in that they were trying to explain certain types of observed behavior, while Maynard Smith and Williams were responding to the general use of group-selection reasoning. Finally, the evidence they used also varied: Wynne-Edwards and Lack largely disagreed about interpretations of observations and field studies, while Maynard Smith and Williams used primarily mathematical and theoretical arguments, respectively.

Approaches to Multilevel Selection

In this section I will describe four approaches to the study of multilevel selection, starting with the goals and questions of each approach and then describing the methods of each approach with some examples of publications that represent those methods.

Theoretical, single level

Goals and questions: This approach includes questions aimed at understanding which level(s) are most important in evolutionary processes, which level should be considered privileged above all others, and whether selection is possible at particular levels such as the group or species. As I described above, the theoretical, single level approach was used extensively in group selection debates during the 1960s and 1970s, including the work of Maynard Smith (1964) and Williams (1966). This is also the

approach of Dawkins's (1976) "gene's eye view," and the approach that Dawkins (1978, 1982) and Hull (1980) had in mind regarding replicator/vehicle and replicator/interactor. Brandon (1982) created a similar distinction, between "unit of selection" and "level of selection," that was also intended to separate those units that provide continuity across generations and the level at which selection is actually acting.

As others have noted (e.g., Lloyd 2001, 2005), the distinctions used by Dawkins, Hull, and Brandon seem to divide into two general questions from the history of group selection debates: 1) What are the units that can serve to reproduce information and pass it from generation to generation – that is, what serves as the unit of heredity (replicators or units of selection) and 2) How does selection act upon phenotypes or levels of organization – that is, what are the vehicles/interactors upon which selection acts, or the levels of selection. One important aspect of this latter question was understanding the dominant level upon which selection acts, such that it made the other levels less important.

The point that many of the early discussions in the history of group selection fall within this approach is easily explained by the fact that there were few experimental studies on the topic prior to the 1970s, and that during the 1960s and 1970s many of the discussions about group selection focused on the feasibility of group-level selection as contrasted with organismal-level selection. Thus, the goal was to describe the most important level at which selection operated rather than understanding the dynamic of multiple levels acting simultaneously.

Methods and examples: Researchers and philosophers using the theoretical, single level approach use largely conceptual methods for thinking about selection,

including the replicator/interactor and the unit/level distinctions. As these were already discussed in chapter 1, I will forgo a detailed discussion of them here. What is relevant for this chapter, however, is that they were intended to help biologists identify and separate the levels that were capable of responding to selection and driving an evolutionary response from those that were important but not the ultimate unit of selection. As such, these were particularly useful during discussions about what kinds of levels of organization might be capable of evolution as a result of natural selection.

Brandon (1982, 1990) proposed another means of identifying the level at which selection is acting, by identifying the level of biological organization that "screens off" other levels from selection in multilevel systems. Essentially, Brandon argued that screening off could be used to determine the actual level that selection is acting upon in a given case from the other potential levels of selection in that case. Following the distinction between phenotype and genotype, Brandon argued that because selection acts on the phenotype rather than the genotype, the phenotypic level will be the level at which selections operates. The way to identify this level, Brandon argued, is to find the level that statistically "screens off" all other levels:

> If A renders B statistically irrelevant with respect to outcome E but not vice versa, then A is a better causal explainer of E than is B. In symbols, A screens off B from E if and only if $P(E,A \square B) = P(E,A) \neq P(E,B)$. (Brandon 1990:83)

This is to say that, when put into terms applicable for a selection experiment, if a trait at level A can statistically account for all of the response to selection in question, while a trait at level B cannot, then it is level A and not B that is the appropriate level of selection for the response observed. Thus, Brandon argued:

> Selection occurs at a given level (within a common selective environment) if and only if 1. There is differential reproduction among the entities at that level; and 2. The 'phenotypes' of the entities at that level screen off properties of entities at every other level from reproductive values at the given level. (Brandon 1990:88)

Brandon also added that his definition for levels of selection was "designed to pick out levels of interaction" (Brandon 1990:93). For this reason, Brandon argued that structured-deme models—that is, IGS models—of group selection can be true group selection models when they identify the level most important in particular cases for the evolutionary process by identifying whether particular traits are adaptations to selection at that level.

Other examples of the theoretical, single level approach include Maynard Smith's (1964) haystack model exploring the conditions under which he expected group-level selection to be a viable force, and Wilson's (1975) trait-group model and the argument that a trait-group could be a level of selection that would not suffer from the criticisms leveled against other group selection models.

Theoretical, multilevel

Goals and questions: This approach includes questions related to how to best conceptualize or operationalize selection so as to allow researchers to investigate the dynamics of selection working across multiple levels of biological organization. In many cases, this includes a particular interest in understanding the net result of selection acting at multiple levels simultaneously, such as when two or more levels exhibit selection acting in opposite directions or cases in which selection at multiple levels might act synergistically.

Methods and examples: Biologists and philosophers using a theoretical, multilevel approach have added to how we might envision selection acting at multiple levels. For instance, Wilson and Wilson (2007) described "selection vectors" in a way analogous to how physicists describe forces acting upon an object. Wilson and Wilson argued that just as we would think of two equal but opposing forces as canceling out, so too can we think of selection at multiple levels of opposite direction but equal intensities as resulting in no response to selection even in the presence of heritable variation. This way of envisioning multiple selection pressures also applies nicely to cases of countervailing selection pressures, in which one force largely negates the effect of the other, as well as cases of complementary selection where the forces combine and potentially result in a larger response than from either alone.

Sober and Wilson (1998) outlined three steps for assessing the dynamics of selection in a multilevel case, though they focused on just two levels of selection: the group-level and the organismal-level. First, "determine what would evolve if group selection were the only evolutionary force" (Sober and Wilson 1998:103). Next, "determine what would evolve if individual [organismal] selection were the only evolutionary force" (Sober and Wilson 1998:103). Finally, "examine the basic ingredients of natural selection at each level" (Sober and Wilson 1998:104); that is, the phenotypic variation, the heritability of that variation, and the differences in survival and reproduction that are caused by that variation.

For other examples of this approach, see Okasha (2006), who provides a sweeping review of the conceptual framework for multilevel selection including theory, tools, and ongoing disputes from both biological and philosophical literature.

It is also worth noting here that Brandon's (1990) "screening off" method could be used instead as a means of testing whether selection in a particular system is working at multiple levels of organization. Though Brandon originally intended it to help identify the particular level at which selecting is working, if knowledge about two or more levels is required to predict the final outcome this would serve as evidence that selection is working synergistically across levels simultaneously.

Empirical, single level

Goals and questions: This approach is best represented by researchers conducting experimental studies, either in the laboratory or the field, focused on questions such as whether a particular level is capable of responding to artificial selection at that level, or whether selection at a particular level is driving the evolution of a trait in question. This approach also includes questions about whether a system is better described as being driven by, for example, group-level selection or organismal-level selection. As such, these questions deal with particular systems or organisms and, though the authors may conclude with possible generalizations of their work, these projects initially begin with experimental or observational questions.

Methods and examples: Statistical methods that provide the ability to judge whether or not a response to group selection was seen in experimental and observational data typify many of these studies. Some of the methods used include analyses of variance (ANOVAs), mixed-effects models, and Monte Carlo simulations (Baer *et al*. 2000; Swenson *et al*. 2000a, 2000b; Lively and Bashey 2009).

Examples of this approach include Goodnight's (1985) experiment using the cress *Arabidopsis* that tested if groups of plants would respond to group-level selection for increased total leaf cover, Baer *et al*.'s (2000) experiment that tested if populations of fish would respond to group-level selection for increased or decreased population size, and Lively and Bashey's (2009) experiment with nematodes that was designed to test if populations of nematodes would respond to group-level selection for increased or decreased population size.

Swenson *et al*. (2000a, 2000b) provided two examples of this approach studying levels other than the group, with separate experiments using unidentified soil communities to test if entire communities would respond to selection for increased or decreased environmental PH, promotion or inhibition of plant growth, or ability to degrade toxic compounds (3-chloroaniline).

Empirical, multilevel

Goals and questions: This approach is used by researchers conducting experimental studies, either in the laboratory or the field, focused on questions related to the interaction of selection acting across levels. This includes studies that explicitly incorporate a component addressing the interaction among two or more levels in relation to the total response to selection. This approach also includes questions about how conflicting selection at two or more levels might result in a lessened response than otherwise expected, or how selection at two or more levels might combine to result in a greater response than would be seen by selection at either level alone.

Methods and examples: The empirical, multilevel approach often relies upon statistical methods that provide the potential for decomposing the total variance of a response to selection into components from multiple levels. For instance, these methods allow a researcher to assess how much of a given response was due to a change at the group-level and how much was due to a change at the organismal-level. The first of these methods was pioneered by George Price, who realized that the mathematical model he created, now called the Price Equation (Price 1970), allowed for the decomposition of variation among generations, and also among levels of organization (Hamilton 1996). Another method growing in popularity is that of contextual analysis, which started as a tool from the social sciences where it was used to explore how a group-level variable could explain variation in individual behaviors (Boyd and Iversen 1979). For example, if considering the voting preferences of individuals, in which individual-level variables such as age or sex have an influence on a particular individual's preferences, the average voting decisions of the neighborhood (= group level) within which the individual lives might also influence his or her ultimate voting decisions (Boyd and Iversen 1979). For a comparison of contextual analysis with the Price Equation, see Okasha (2004b, 2006).

Another example of this approach is Wade's (1976, 1977) work on group-level selection in flour beetles. Wade included a treatment and control to allow for analysis of two conflicting levels by setting organismal-level selection for increased population size against group-level selection for decreased population size, as well as complementary levels of selection by setting organismal-level selection for increased population size against group-level selection for increased population size. Goodnight *et al*. (1992) provide an example of how the same researcher(s) can use more than one approach, as

Goodnight *et al*. (1992) reanalyzed Goodnight's (1985) data in order to illustrate the benefits of contextual analysis. Another example of mechanistic explanations for group-level selection is Amdam and Page's (2010) work on honeybees. They describe the results of an ongoing series of studies that started by selecting on beehives for increased and decreased pollen storage, including the follow-up studies that tracked the responses at the level of the hive, as well as the behavioral, physiological, and developmental changes in the foraging females that produced the hive-level responses (see also, Page and Amdam 2007; Page and Fondrk 1995)

Others

The four approaches discussed above are the extreme examples created by the intersection of the two continua I have described. Many researchers work somewhere between the two ends of theoretical and empirical, but remained interested in a single level of analysis. For example, Wynne-Edwards (1964) and Lack (1966) were both interested in theory creation but were motivated by, and heavily grounded in observations. A similar example is Damuth and Heisler (1987), who advocated the use of contextual analysis in multilevel selection, despite the fact that they did not conduct novel experimental studies themselves.

Another kind of research that falls somewhere between the approaches I described above is that of researchers who largely rely upon agent based models for multilevel research. Some of these fall within the focal level end of the continuum, such as those exploring how organismal-level rules of behavior can result in emergent and heritable properties of the group (for examples, see Miller and Page 2007). Others fall more to the

multilevel end of the continuum by using agent based models to explore the dynamics of selection acting upon multiple traits at multiple levels (for examples, see Miller and Page 2007). Indeed, in recent years some modelers have even taken to recreating, in agent based models, some of the disputed verbal models of the past such as Wynne-Edwards's (1962) thesis described in chapter 2. Among these, Werfel and Bar-Yam (2004) demonstrated that there are conditions under which the group selection envisioned by Wynne-Edwards (1962) can indeed result in the maintenance of group-beneficial traits, a conclusion also supported by experimental evidence (e.g., Kerr *et al*. 2006). Agent based models have also been used by researchers studying human behavior, including how particular patterns of individual decision making can result in group-beneficial outcomes (see Boyd and Richerson 2005; Henrich and Henrich 2007).

Interpretation of Experimental Data

As a demonstration of the conflicting interpretations that can result from the different approaches described above, I now turn to a description of a single experiment and the potential interpretations that arise from each of the approaches.

Overview of the *Nasonia* experiment in chapter 1: Recall that the group selection treatment included a collective context of five wasps housed together followed by a solitary context of each wasp housed separately. The organismal selection treatment included a solitary context in which each wasp was housed separately followed by a collective context in which the wasps were placed in groups of five.

In the group selection treatment, the group context did not show a response to selection for increased population size, while the solitary context did show a response.

Within the organismal selection treatment, there was a response to selection in the solitary context, but not in the group context.

Theoretical, single level interpretation: The theoretical, single level approach might rely upon "screening off" to identify whether group-level selection was at work in the *Nasonia* experiment, which would require that population size at the group-level screened off the fecundity at the organismal level. However, in describing how to apply screening off, Brandon (1990) explicitly argued that population size was an inappropriate group-level character because it should be interpreted as an aggregate of individual fitness. Thus, based on Brandon's description, the *Nasonia* experiment would be interpreted as simply showing organismal selection in two separate environmental contexts: the solitary context and the collective context.

Another possible interpretation from the theoretical, single level approach might be to conclude that because the groups in the *Nasonia* experiment did not respond to what I called group-level selection, but the females in the solitary context of the group-selection treatment did, it is possible that what I called group-level selection was actually indirect selection for an organismal trait (Okasha 2006) or what Sober (1984) would call "selection for" as opposed to selection "selection of." That is, the trait that I thought I was selecting upon at the group level was actually not directly connected to organismal fitness, but was correlated with a trait at the organismal level that did respond to the selection treatment.

Theoretical, multilevel interpretation: This approach is largely concerned with how to conceptualize selection operating at multiple levels, so there are few analytical tools that apply to empirical cases. Those that have been mentioned or advocated are

usually contextual analysis or the Price equation. Unfortunately, because the organismal fitness proxy within the group selection treatment failed to provide appropriate proxies for organismal contribution to group fitness, the contextual analysis component was removed from the experiment. This also means that the Price equation would not be appropriate for use in the analysis, and so the theoretical, multilevel approach would not be able to apply either of these analytical tools. Instead, a researcher using this approach might rely upon the general description provided by Sober and Wilson (1998) to guide the analysis, though because this method was intended for approaching an adaptation with no prior knowledge of the selection processes that led to its creation, its application to experimental data might be awkward.

It is also important to note that many researchers using this approach advocate a plurality of conceptual tools, such as both multilevel selection and inclusive fitness models. This leaves the possibility of a plurality of interpretations each being considered for a single set of data. For instance, such a plurality might include the single level interpretation that the wasps in the experiment are responding to the presence of other females and adjusting their clutch sizes to maximize organismal fitness—an argument similar to Lack's (1966) interpretations—while also accepting a multilevel interpretation that groups of wasps do not interact in ways that lead to a heritable group-level trait for population size.

In regard to this second point, this experiment could be viewed as consistent with Wilson's trait group models (e.g. Wilson 1975), in which the interactions of the individuals within the collective vials make them valid groups for the trait of population size. This would mean that the potential for selection was present even if there were no

response to population size observed as a result. However, such a view would use the experimental data to track the influence of group membership on organismal fitness in an MLS1 analysis, whereas the analysis of the *Nasonia* experiment described in chapter 1 focused on an MLS2 analysis, tracking the fitness of entire groups.

Empirical, single level interpretation: This approach generally is used to test if an experimental treatment responded to group-level selection. Thus, this approach would accept or reject the *Nasonia* experimental data as demonstrating a response to group selection based on the statistical significance of the results. An important point to note here is that both empirical approaches would accept that group selection was present if it were part of the experimental design. The experiment would then provide the data necessary to determine if there were a statistical response to the artificial selection that the experiment imposed.

In the case of the *Nasonia* data, the experimental design was intended to create circumstances favoring group-level selection for increased number of offspring per standardized host. However, an analysis of the data did not detect a statistically significant difference between the initial number of offspring per host and the number of offspring per host after the experiment. Thus, this approach would conclude that group-level selection was present in the experimental treatment, but that it did not lead to a response under the conditions of the experiment.

Empirical, multilevel interpretation: This empirical approach would also accept that group-level selection was present if the experiment were designed to exert selection at the group level. In the *Nasonia* experiment, this would be a treatment for group selection and a treatment for organismal selection. Rather than focusing on whether

selection was present, the interest of this approach would be in determining the response, if any, and how much of that response might be attributed to selection acting at different levels.

The *Nasonia* experiment was designed with a component that would have allowed for the use of contextual analysis by providing a proxy for organismal fitness despite the fact that the fitness of each wasp was unavailable for this experiment. This proxy is important because contextual analysis requires both organismal and group fitness in order to decompose the variance. However, the contextual analysis component of the *Nasonia* experiment was subsequently dropped because the wasps behaved in very different ways in the proxy condition than they did in the group conditions, making the proxy a poor indicator of actual organismal contribution to group fitness.

Despite the lack of contextual analysis, researchers using the data driven, multilevel approach would likely notice the variations observed in the group selection treatment. In particular, there was a statistically significant difference in number of offspring per host between the beginning and ending of the experiment within the group treatment when the females were moved into the solitary context so that each had a vial separate from other wasps. In other words, despite the lack of response to group-level selection when the wasps were housed in groups, one interpretation is that there was a response to group-level selection on the wasps when they were then transferred to solitary context vials.

Conflicting Interpretations

In this section I will first discuss some conflicting interpretations that can arise between the approaches described above. Then, I turn to conflicts that can exist even within the same approach. Finally, I address some of the causes for these conflicts, including differences in both methods and epistemology.

Conflicting interpretations between approaches: Even with the very same experimental design and results, the different approaches I have described might well lead researchers to different interpretations of experimental data. Two of the most extreme differences are in the theoretical, single level and in the empirical, multilevel.

As I described above, the theoretical, single level approach supports an interpretation of the *Nasonia* experiment as demonstrating that no group selection had occurred, and some might use such a conclusion as evidence that group-level selection is not an important force for evolution. Indeed, it is worth mentioning again that Brandon (1990) argued that "population size" was not a legitimate group-level trait because the size of a population would only be an aggregate of the organisms' offspring. This, of course is in direct contrast with many of the experimental studies of group-level selection, including those from both the empirical, single level approach and the empirical, multilevel approach. In many of those studies, population size is taken to be an emergent trait that is only present within an entire population because no single individual can be said to possess the character "population size." Additionally, in any case where individuals change their fitness-relevant behavior as a result of group membership, the total population size will be an emergent character of the interactions within the population. It is also interesting to note that, despite Brandon's (1990)

argument that population size is not a legitimate group-level trait, Brandon (1990) highlighted Wade's (1977) work as representing an experimental demonstration of group-level selection.

Furthermore, some researchers using the theoretical, single level approach might argue that because the groups in the *Nasonia* experiment did not respond to group-level selection, but the number of offspring per host of the solitary wasps did change in both treatments, the experiment is actually a demonstration of the importance of organismal-level selection in various contexts. They might add to this that the experiment also demonstrated that the apparent group-level traits under study were really only the result of indirect selection on organismal traits.

However, researchers from the empirical, multilevel approach—while, I presume, disappointed that the experiment did not successfully include a covariance analysis—might still interpret the change in number of offspring per host within the solitary context of the group selection treatment as representing a response at the organismal-level to group-level selection. Such a conclusion would be in contrast with the single level approaches, which often stipulate that responses and adaptations to selection can only occur at the level at which selection is acting.

Recall that the types of questions asked by the single level approaches are often concerned with the evolution of a particular trait. Indeed, some of the arguments from the theoretical, single level approach are built around a researcher identifying *the* particular trait that selection is acting upon before constructing an explanation for why that trait is an adaptation. But many of the multilevel studies involve multiple traits spread across two or more levels of organization because the interest is in understanding selection

acting at multiple levels simultaneously. As such, a theoretical, single level researcher might find the interpretations of an empirical, data driven researcher to be entirely unsatisfactory. For example, Lieberman and Verba (1995) argued:

> Wade [1977] and Wilson [1977] identified cases of group selection that we classify as cases of group sorting and/or context dependence, because in their examples the sorting of groups was either caused by properties of organisms or the groups were not even sorted but organisms were affected by their group membership (Mustapha Mond sorting, according to Vrba 1989). Cases of group sorting only equate with group selection if there are emergent group characters and if their presence leads to differential group birth and/or death rates. (Lieberman and Vrba 1995:395)

In other words, because these experiments included responses in traits at multiple levels, Lieberman and Vrba did not consider them to be representing true group-level selection. Also note that Lieberman and Vrba are implicitly arguing that the only group-level selection that they accept is that of MLS2, since they require differences in birth or death rates of the groups.

As another example of conflict caused by different approaches with respect to the level that is the focus of the research, consider the debate discussed in chapter 2 between West *et al*. (2007, 2008) and Wilson (2008). Though I used Wilson's (1975) early discussions of trait-group models as an example of the theoretical, single level approach, much of Wilson's later work fits the theoretical, multilevel approach because the questions moved from whether trait-groups can serve as a level of selection to how selection acting at the organismal level as well as the trait-group level might simultaneously influence evolution of organisms. Thus, Wilson (2008) seems most interested in questions about how the group membership can influence organismal fitness. In contrast, West *et al*. (2007, 2008), advocating kin selection, seem more

interested in questions that fit with a theoretical, single level approach. Their arguments in favor of kin selection suggest that their questions and goals are focused on the organismal level and how behavioral traits can influence related organisms.

In summary, many conflicts can arise between these different approaches, including conflicting interpretations of the same data. Indeed, these differences can even lead to cases where the same data are used by opponents in a single debate, with each arguing that the data support his or her own position (e.g., Wynne-Edwards 1962; Lack 1954, 1966). The differences in interpretation among these approaches are further exacerbated by the fact that the type of evidence that one approach finds strongly supporting an argument might to another approach seem completely irrelevant. These two points suggest an explanation for why the addition of new data has not resolved the overall controversy or particular conflicts.

Conflicting interpretations within approaches: Even the approaches that I have described above do not fully capture the nuance of interpreting data and I fully expect that researchers using the same approach might reach different interpretations contingent upon how their approach intersects with their interests in MLS1 or MLS2. As another example of how approaches can influence data interpretation, reconsider the pair of similar experiments that I mentioned in chapter 1: Baer *et al*. (2000) and Bashey and Lively (2009).

Baer *et al*. (2000) selected on total population size with the live bearing fish *Heterandria formosa*, in which neither source population showed a statistically significant response in either increased or decreased population treatments. Specifically, Baer *et al*. started with two source populations and two selection treatments (high and

low) for population size. They placed 15 males and 15 females in each of four replicate tanks for each source population and treatment, and then counted the census size after six weeks. At this point, they identified the two tanks with the highest (or lowest) populations and then equally distributed those fish into the four tanks of the next "generation" in proportion to their class: males, females, and juveniles. For example, if the two highest tanks contained 8 males and 10 males respectively, then they would distribute 4 males into each of the next generation's tanks. After six rounds of selection—or what they called six generations—Baer *et al*. counted the final population sizes and compared these with the initial population sizes, concluding that there was no significant response to selection in any of the populations.

Bashey and Lively (2009), studying the endopathic nematode *Steinernema carpocapsae*, selected on increased emergence of nematodes per host. They ran a similar experiment with three populations derived from a single mixed stock, and of the three populations only one showed a response to selection on the group level character of number of nematodes emerged from the host after 14 days. Specifically, Bashey and Lively performed selection on increased and decreased emergence rate from hosts, using five replicate lines for each treatment type in each of the three experiments they performed, and each line consisted of 20 hosts. The life cycle of parasitic nematodes involves multiple adults infecting a single host, mating inside the host, and then a single generation of offspring emerge. Thus, host caterpillars were infected with approximately 15 nematodes in 0.5mL of deionized water. Bashey and Lively estimated that under the conditions of the experiment, approximately 50% of these nematodes survived to reproduce within the host. After 7-14 days, the offspring began to emerge from the hosts

and after 14 days the number of nematodes were counted and then selection was performed based on the number of emerged nematodes standardized by host mass. These nematodes were then used to infect the next 20 hosts of that line. This was repeated for 4 rounds of selection. Bashey and Lively then compared the number of nematodes emerging in the high-line with those in the low-line of each experiment, and determined that only in experiment 1 were the two treatments significantly different, concluding that there was a response to selection only in the population used for experiment 1.

Notice that of these two experiments, the nematode experiment is the most similar to the *Nasonia* experiment, partly as a result of each being built around a parasitic species. Even so, both the fish and the nematode experiments included high and low lines, selecting for increased and decreased population sizes respectively, as well as relying upon differential extinction of groups to create the selection gradient.

However, the Baer *et al*. (2000) study differed from the Bashey and Lively's (2009) nematode experiment—and the *Nasonia* experiment—in key respects. For instance, as mentioned in chapter 1, both the nematode experiments and my *Nasonia* experiments used propagule pools for creating subsequent generations while the fish experiment used migrant pools. This means that for each generation in the *Nasonia* experiment the largest population of five was the only one that created offspring, and in the nematode experiment it was the largest of 20, whereas in the fish experiment the experimenters merged offspring from the two largest of the four tanks. As mentioned previously, migrant pools disrupt the population structure, but this method for selection also resulted in a lower selection gradient since only fifty percent of the populations went

extinct without contributing to the next generation as compared with eighty percent in the *Nasonia* experiment.

The differences between the use of MLS1 and MLS2 are also implicit in these accounts. Bashey and Lively (2009) used MLS2, demonstrated by their interest in how entire groups responded. Baer *et al*. (2000) were interested in MLS1, as is clear from the way that Baer *et al*.'s experiment focused on the fish used in the groups, rather than the group properties themselves. For instance, the fish in a selected group were moved from one "generation" to another, leading Baer *et al*. to call their experiment one with "overlapping generations." However, notice that the group generations did not overlap. Rather, the overlap of generations only applied to the organisms within the groups. Thus, in order to be considered overlapping generations, the interest must have been on the fish themselves rather than the groups of fish. This interest in the organismal-level is also clear in their choice of a migrant pool rather than a propagule pool so as to avoid inbreeding, as mentioned in chapter 1.

Finally, the conclusions of these two experiments differed greatly, which was partly influenced by the shared approaches, but was also influenced by differences in the aims of the experiments and what they were tracking—that is, group membership influencing organismal fitness *sensu* MLS1 or group fitness via group reproduction *sensu* MLS2—despite the fact that they used the same approach. Bashey and Lively (2009) found that two of the three experiments did not respond to group-level selection, but in their conclusion they focused on the possible explanations for why the one population responded when the other two did not. Thus, their tone was supportive of group-level selection as a process, and understanding the conditions under which group-level

selection might lead to a response. In contrast, Baer *et al*. (2000) did not find a response and concluded that group-level selection must necessarily be a weak force that would rarely occur. They did not discuss possible explanations, including the possibility that the conditions of their experiment accounted for their results and conclusions, as the design was not favorable for the type of between-group variation that group-level selection is most likely to act upon.

Conflicting methodology and epistemology: In addition to the conflicting interpretations described above, conflicts can also arise as a result of the particular concepts, definitions, and assumptions used by different approaches. For example, consider Maynard Smith's (1964) haystack model that was designed as a critique of Wynne-Edwards's (1962) group selection thesis. Maynard Smith used the restrictive conditions under which the haystack model could work to argue that the conditions necessary for group-level selection to occur were so rare that it would never be an important force in nature. But Wynne-Edwards (1964) replied with an argument that Maynard Smith's model was too abstract:

> The model of the mice in the haystacks is not, perhaps, a sufficiently close approximation to any natural situation to help us far towards a solution. A realistic counterpart might be, for example, the woodlice (*Porcellio scaber*) that fed on the green alga *Protococcus* living on tree-trunks studied by Brereton [1957]; marked woodlice confined their feeding to their own particular tree, and the population appeared to be subdivided thus into breeding units. Had any of the latter increased too freely they could have exterminated their stock of this particular food plant, which does not regenerate easily. (Wynne-Edwards 1964:1147)

Notice that, in essence, Wynne-Edwards disagreed with the assumptions built into the way that Maynard Smith's haystack model defined "group" and argued instead for a more biologically realistic example, in which case he argued that his (1962) thesis of

group selection as a result of differential extinction could indeed work in natural populations.

However, Wynne-Edwards (1964) also acknowledged that there was a deeper conflict between his views and those of Maynard Smith:

> The major obstacle to the constructive discussion between us really arises from the understandable (though regrettable) differences in outlook and experiences between a laboratory geneticist and a field ecologist. To me his picture of the territorial systems and other aspects of conventional behaviour appears scarcely true or comprehensive enough to provide a basis for valid deduction; my own grasp of the genetical theory of natural selection, on the other hand, no doubt looks still more halting and inept to him. We ought to enlarge the area of common ground, but that is too big a task to discharge effectively here. (Wynne-Edwards 1964:1147).

This difference in "outlook and experiences" suggests that Wynne-Edwards was really highlighting the fact that he and Maynard Smith disagreed not only on the specific definitions and assumptions used in their models but also on deeper epistemological issues at stake.

Wynne-Edwards was not the only one to express such concerns about the definitions, assumptions, or concepts used by others. Maynard Smith (1998:217) shared in an interview his own complaints regarding the group selection discussions of the 1970s and 1980s:

> Hamilton, myself, Dawkins and so on were all thinking in terms of gene frequencies and the sort of Haldane-Fisher type models that we were used to. The people in the States who took up the problem, tended to formulate it in terms of variance-covariance matrices and things of this kind, much more analogous to classical thermodynamics. You couldn't see the genes for the variances. I have to say that I find this stuff impenetrable, in much the same way that I find classical thermodynamics impenetrable. I can use it for working out a steam engine's behaviour but I don't understand it; it doesn't give me an insight into what I think is going to happen next. (Maynard Smith 1998:217)

Notice here that Maynard Smith's argument is not that the co-variance methods were inadequate or improper. Rather, he argued that they were not useful to his way of understanding the world. Of course, Maynard Smith disagreed about more than simply the co-variance methods used by the "new" researchers during the late 1970s and 1980s:

> By 1975, individual selection had become the orthodoxy, and a lapse into group selection was liable to be met with hatred, ridicule and contempt. In science, however, it seems that the establishment of an orthodoxy is the signal for attempts to undermine it. These have come from two directions—attempts to re-establish the role of selection at levels above the individual, and attempts to replace individual by gene selection. (Maynard Smith, 1998:206)

> Unfortunately, they [group selection researchers after 1975] used all the words we had been using but in totally different senses. So the term 'group selection' that had a perfectly clear meaning in the debates in Britain between 1961 and 1975, was used in a completely different sense between 1975 and 1985 by young people in America, which was a damn nuisance. We had to go over all sorts of old vomit again. (Maynard Smith 1998:217)

Thus, Maynard Smith had two distinct complaints about the group selection discussion of the late 1970s and early 1990s: 1) he disliked the application of covariance equations to the discussions of group selection and 2) he considered the debate settled until "young people in America" started it again using new definitions for existing terms – specifically, as, seen in chapters 1 and 2, the new definition of "group" used by intradermic group selection models.

Of course, as I suggest in chapter 1 and discuss in chapter 2, even during the times that Maynard Smith mentioned there were different usages of the "perfectly clear" meaning for words like "group." Some of these confusions were not sorted out until the late 1980s—such as the difference between group selection of the MLS1 or MLS2

varieties described by Heisler and Damuth (1988)—while others, like variations in the definition of "group," remain even now.

Equally, conflicts caused by different worldviews have remained. For instance, Dawkins (1994), in responding to Sober and Wilson (1994) wrote:

> Wilson & Sober's passion is obviously genuine. I welcome their plainly sincere attempt at clarification and, despite myself, I quite enjoy the rhetoric. They are zealots, baffled by the failure of the rest of us to agree with them. I can sympathize: I remain reciprocally baffled by what I see as the sheer, wanton, head-in-bag perversity of the position that they champion. (Dawkins 1994:616-617)

Indeed, this observation seems to apply to other disputes over multilevel selection, including the disagreement described earlier between West *et al*. (2006, 2008) and Wilson (2008). Again, the heart of these arguments is not experimental data or particular interpretations; these disputes are about how satisfactory the concept of multilevel selection is when compared with the concept of inclusive fitness. In this regard, the opponents of multilevel selection, including Dawkins and West *et al*. seem to share Maynard Smith's views that multilevel selection does not provide an advantage for understanding natural selection:

> The question of whether there is a group selection or not is a question of the nature of the world. Whether or not you should use inclusive fitness or gene-centered models is not a question of what the world is like, it's a question about what is a convenient way of modeling it. (Maynard Smith, 1998:214)

Conclusion

The role of data: Brandon (1994), discussing the relationship between theory and experiments, argued:

> Wade's [1977] laboratory demonstrations of the effectiveness of group selection were extremely important and are certainly to be commended. But they have only gone a short way toward resolving the controversy concerning group selection; they have not convinced any skeptics of the importance of group selection in nature. For that, one needs experiments under more 'natural' conditions: one needs field work. (Brandon 1994:68)

I agree with Brandon that the experimental data seem to have done little to change the views of those who oppose multilevel selection. However, given what I have described above regarding the alternative interpretations of laboratory experiments resulting from different approaches to multilevel selection, I would argue that the addition of new studies—field-based or otherwise—will affect the ongoing theoretical arguments little more than those that have already been published.

The reason for this lies in the nature of the conflicts at the heart of these disputes. Okasha (2006) captured an important part of this relationship between data and theory when he wrote in his review of multilevel selection:

> Obviously, empirical data is crucial for resolving the levels-of-selection question, as for all scientific questions; but conceptual clarity is a prerequisite too. Unless we can agree on what it means for there to be selection at a given hierarchical level, on what the criteria for individuating 'levels' are, on whether selection at one level can ever be 'reduced' to selection at another, on how multi-level selection should be modeled, and on whether there is always 'one true fact' about the level(s) at which selection is acting, then there is little prospect of empirical resolution, however much data we collect. (Okasha 2006:2)

I argue that most of the debates regarding both group selection and multilevel selection are not about data. Instead, they are the product of conflicting views of, as Maynard Smith called it, the "nature of the world" specifically, the evolutionary process and what entities are subject to it. To this end, data are and will continue to be essential. But it is not data alone that the debates of the past or present hinge upon, so the addition of new

data will not in and of itself help to resolve the controversy. Instead, dialogue about researchers' views regarding the nature of the world might be necessary in addition to adding more biologically relevant data.

Approaches and debates: In chapter 2 I argued that the history of group selection, and now multilevel selection, is best viewed as a series of interrelated debates about specific topics rather than as a single continuous history. Here, I have described multiple approaches to the study of multilevel selection that also help account for the ongoing controversy. Although these approaches share a general subject, understanding multilevel selection, the particular ways that researchers frame their questions, the general methods that they use, and even the definitions of the concepts that they depend upon can vary from approach to approach, leading to the potential for quite varied interpretations with respect to both data and theory. However, the conflicts resulting from the different approaches are not insurmountable. By recognizing that different approaches can have different explanatory goals, researchers might not judge the work of others with the assumption that what they seek to explain is the same. Instead, theoreticians, philosophers and experimentalists should all describe what they take their work to assist in explaining. Finally, when necessary, researchers should give greater voice to the epistemology motivating their work rather than only discussing the theory and data that are built upon the—often-unmentioned—worldviews.

CONCLUSION

The notion that natural selection might act at levels of biological organization other than only those of organisms or their genes has a long history including both theoretical and empirical analysis. But controversy remains despite more than six decades of study and I have explored in several ways why this controversy has continued for one particular case of multilevel selection: groups as a level at which natural selection may or may not operate effectively. In doing so, my research illustrated the role that conceptual change has played in the controversy about group-level selection over the past sixty years. I argue that conceptual clarity as well as empirical evidence are necessary to resolve the controversy.

The experimental work described in the first chapter demonstrated that the interpretation of experimental results can be confounded by the multiplicity of implicit definitions for the term "group." The historical analysis in the second chapter provided a review of modern multilevel selection theory by demonstrating the ways in which the debates about levels of selection have and have not changed. I also showed that the aspects of the debates that have changed the most are those related to the theoretical content and empirical data available, while disputes related to methods remain largely unchanged. In the second chapter I argued that the history of the group selection controversy is best interpreted as multiple, interrelated debates rather than a single continuous debate. Finally, my synthetic analysis built upon the experimental and historical work to provide a framework for understanding the ongoing disputes about group selection, and multilevel selection more generally. In particular, I argued that there are different "approaches" to the study of multilevel selection distinguished by where the

research falls upon two axes: 1) whether the researcher is primarily interested in a single level of organization or multiple levels, and 2) the extent to which the work starts from a theoretical or empirical perspective. I described four such "approaches" and discussed how these approaches might lead to conflicting interpretations of the *Nasonia* experimental data from chapter one.

The lessons to be learned from this project are varied, as are the potential for avenues of future research. Indeed, there are lessons to be learned not only for those studying the topic of multilevel selection, but also lessons for historians of biology and lessons that might apply broadly to controversies in the sciences beyond the biological.

For the empiricists and theoreticians studying multilevel selection this project brings to light the often undiscussed challenges involved in designing experiments to address questions within the subject of multilevel selection. These challenges partly arise from the fact that the theoretical arguments for how to test multilevel selection cases are rarely presented in a way that is easily applied to an experimental system, leaving a separation between the theoretical and experimental explorations of the subject. One of the clear consequences, and challenges, is that what counts as a "group" varies conceptually among investigators; yet, researchers often take the definition that they use to be self-evident. Thus, future work should focus on providing conceptual clarity, while also integrating the experimental and theoretical work that has been done within the subject of multilevel selection. It is also worth reiterating here that continuing research on multilevel selection—in particular the two questions of 1) under what circumstances we should expect to see selection operating at one level of selection or another, and 2) how selection working across multiple levels of organization simultaneously will result in a

single response—will continue to be of importance beyond biology because these biological concepts are used as models for understanding selection acting on complex adaptive systems both biological and otherwise. Indeed, in the future both biologists and philosophers of biology could add much to the growing science studying complex adaptive systems. Equally, there might be generalizable lessons learned from this work that would continue to benefit work on biological multilevel selection, as I believe is exemplified by the work of some anthropologists in using multilevel selection, and group-level selection in particular, to understand cultural evolution (e.g., Boyd and Richerson 1985, 2005; Richerson and Boyd 2005; Henrich and Henrich 2007).

For historians of biology, I have provided some new material, as despite the fact that the study of multilevel selection has been named as such since the late 1980s, to my knowledge there has not yet been a comprehensive review/history of the ideas behind multilevel selection. Histories and reviews of group-level selection capture some of the research and ideas leading to what became multilevel selection, yet fail to capture important elements such as the work on whether species could serve as levels of selection and discussions of community-level selection. Detailed histories exploring the extent to which these subjects overlapped before the coining of the phrase "multilevel selection" would benefit future work by providing the grounding out of which the current multilevel selection questions arose and by further illustrating the role of conceptual change in these discussions. I also hope that my work has provided another example of the usefulness of tracking conceptual change in order to understand both the current state of a subject, as well as understanding any controversy associated with the subject.

Finally, it is this last suggestion that might well apply to fields of research beyond the biological sciences. If tracking the role of conceptual change in promoting or exacerbating ongoing controversy within a field of science provides some means of better understanding that controversy, and thereby suggesting avenues by which the controversy might be alleviated, then this project could serve as one such example of how this might be applied to other areas of research. In particular, this project suggests that other controversies might well be continued as a result of changes in the ways that particular concepts are operationally defined or as a consequence of researchers holding, what Maynard Smith referred to as, conflicting "world views." In either of these cases, it is only by acknowledging the conflicting definitions or world views and then providing operational definitions—in effect, adding conceptual clarity—that such controversy will be resolved.

REFERENCES

Abbot, P., J. Abe, J. Alcock, S. Alizon, J.A.C. Alpedrinha, M. Andersson, J. Andre, M. van Baalen, F. Balloux, S. Balshine, N. Barton, L.W. Beukeboom, J.M. Biernaskie, T. Bilde, G. Borgia, M. Breed, S. Brown, R. Bshary, A. Buckling, N.T. Burley, M.N. Burton-Chellew, M.A. Cant, M. Chapuisat, E.L. Charnov, T. Clutton-Brock, A. Cockburn, B.J. Cole, N. Colegrave, L. Cosmides, I.D. Couzin, J.A. Coyne, S. Creel, B. Crespi, R.L. Curry, S.R.X. Dall, T. Day, J.L. Dickinson, L.A. Dugatkin, C. E. Mouden, S.T. Emlen, J. Evans, R. Ferriere, J. Field, S. Foitzik, K. Foster, W.A. Foster, C.W. Fox, J. Gadau, S. Gandon, A. Gardner, M.G. Gardner, T. Getty, M.A.D. Goodisman, A. Grafen, R. Grosberg, C.M. Grozinger, P. Gouyon, D. Gwynne, P.H. Harvey, B.J. Hatchwell, J. Heinze, H. Helantera, K.R. Helms, K. Hill, N. Jiricny, R.A. Johnstone, A. Kacelnik, E.T. Kiers, H. Kokko, J. Komdeur, J. Korb, D. Kronauer, R. Kümmerli, L. Lehmann, T.A. Linksvayer, S. Lion, B. Lyon, J.A.R. Marshall, R. McElreath, Y. Michalakis, R.E. Michod, D. Mock, T. Monnin, R. Montgomerie, A.J. Moore, U.G. Mueller, R. Noë, S. Okasha, P. Pamilo, G.A. Parker, J.S. Pedersen, I. Pen, D. Pfennig, D.C. Queller, D.J. Rankin, S.E. Reece, H.K. Reeve, M. Reuter, G. Roberts, S.K.A. Robson, D. Roze, F. Rousset, O. Rueppell, J.L. Sachs, L. Santorelli, P. Schmid-Hempel, M.P. Schwarz, T. Scott-Phillips, J. Shellmann-Sherman, P.W. Sherman, D.M. Shuker, J. Smith, J.C. Spagna, B. Strassmann, A.V. Suarez, L. Sundström, M. Taborsky, P. Taylor, G. Thompson, J. Tooby, N.D. Tsutsui, K. Tsuji, S. Turillazzi, F. Úbeda, E.L. Vargo, B. Voelkl, T. Wenseleers, S.A. West, M.J. West-Eberhard, D.F. Westneat, D.C. Wiernasz, G. Wild, R. Wrangham, A.J. Young, D.W. Zeh, J.A. Zeh, and A. Zink. 2011. Inclusive fitness theory and eusociality. *Nature*. 471:E1–E4.

Alexander, R. D., and G. Borgia. 1978. Group selection, altruism, and the levels of selection of life. *Annual Review of Ecology, Evolution and Systematics*. 9:449-74.

Amdam, G.V., and R.E. Page. 2010. The developmental genetics and physiology of honeybee. *Animal Behaviour*. 79:973-980.

Arnold, A.J., and K. Fristrup. 1982. The Theory of Evolution by Natural Selection – A Hierarchical Expansion. *Paleobiology*. 8(2):113-129.

Baer, C. F., J. Travis, and K. Higgins. 2000. Experimental evolution in *Heterandria formosa*, a livebearing fish: group selection on population size. *Genetics Research*. 76:169–178.

Bashey, F., and C.M. Lively. 2009. Group Selection On Population Size Affects Life-History Patterns In The Entomopathogenic Nematode *Steinernema Carpocapsae*. *Evolution*. 63-5: 1301–1311.

Beukeboom L, C. Desplane. 2003. Nasonia. *Current Biology*. 13(22):R860.

Beukeboom, L.W., and L. Van de Zande. 2010. Genetics of sex determination in the haplodiploid wasp Nasonia vitripennis (Hymenoptera: Chalcidoidea). *Journal of Genetics*. 89:333-339.

Bijma P., and M. Wade. 2008. The Joint Effects of Kin, Multilevel Selection and Indirect Genetic Effects on Response to Genetic Selection. *Journal of Evolutionary Biology*. 21(5): 1178–1188.

Borrello, M.E. 2010. *Evolutionary Restraints: The Contentious History of Group Selection*. University of Chicago Press, Chicago.

Boyd, L.H., and Iversen G.R. 1979. *Contextual Analysis: Concepts and Statistical Techniques*. Wadsworth, Belmont CA.

Boyd, R. and P.J. Richerson. 1985. *Culture and the Evolutionary Process*. University of Chicago Press, Chicago.

Boyd, R. and P.J. Richerson. 2005. *The Origin and Evolution of Cultures*. Oxford University Press, Oxford.

Brandon, R.N. 1982. Levels of Selection. *Proceedings of the Philosophy of Science Association*. 1:315-323.

Brandon, R.N. 1990. Adaptation and Environment. Princeton University Press, 78-133.

Brandon, R.N. 1994. Theory and Experiment in Evolutionary Biology. *Synthese*. 99:59-73.

Brandon, R.N. 1999. The Units of Selection Revisited: The Modules of Selection. *Biology and Philosophy*. 14:167-180.

Brenden, F., and M.J. Wade. 1989. Selection within and between kin groups of the imported willow leaf beetle. *The American Naturalist*. 134(1):35-50.

Brucker, R.M., and S.R Bordenstein. 2013. The Hologenomic Basis of Speciation: Gut Bacteria Cause Hybrid Lethality in the Genus Nasonia. *Sciencexpress*. July 18: 1-5.

Burton-Chellew, M.N., T. Koevoets, B.K. Grillenberger, E.M. Sykes, S.L. Underwood, K. Bijlsma, J. Gadau, L. Van de Zande, L.W. Beukeboom, S.A. West, and D.M. Shuker. 2008. Facultative Sex Ratio Adjustment in Natural Populations of Wasps: Cues of Local Mate Competition and the Precision of Adaptation. *The American Naturalist*. 172(3):393-404.

Cassidy, J. 1981. Ambiguities and Pragmatic Factors in the Units of Selection Controversy. *Philosophy of Science*. 48(1):95-111.

Charlesworth, B. 1981. Review of Natural Selection of Populations and Communities, by D.S. Wilson. *Journal of Applied Ecology*. 18:328-30.

Charnov, E. L., Krebs, J. R. 1975. The evolution of alarm calls: altruism or manipulation? *The American Naturalist*. 109:107-12.

Cohen, D., Eshel, I. 1976. On the founder effect and the evolution of altruistic traits. *Theoretical Population Biology*. 10:276-302.

Collins, J.P. 1986. "Evolutionary Ecology" and the Use of Natural Selection in Ecological Theory. *Journal of the History of Biology*. 19(2):257-288.

Collins, J.P. 2003. What can we learn from community genetics? *Ecology*. 84:574-577.

Coyne, J.A. 2012. The Demise of Group Selection. "Why Evolution is True" website. http://whyevolutionistrue.wordpress.com/2012/06/24/the-demise-of-group-selection/ Accessed 9/5/2013.

Craig, D. M. 1982. Group selection versus individual selection: an experimental analysis. *Evolution*. 36:271–282.

Craig, J. V., and W. M. Muir. 1996. Group selection for adaptation to multiple-hen cages: beak-related mortality, feathering, and body weight responses. *Poultry Science*. 75:294–302.

Damuth, J. and I.L. Heisler. 1988. Alternative Formulations of Multilevel Selection. *Biology and Philosophy*. 3:407-430.

Darwin, C. 1871. *The Descent of Man, and Selection in relation to Sex*. Republished 1981 Princeton University Press, Princeton NJ.

Dawkins, R. 1976. *The Self Gene*. Oxford University Press, Oxford.

Dawkins, R. 1979. Twelve misunderstandings of kin selection. *Zeitschrift Fur Tierpsychologie*. 51:184-200.

Dawkins, R. 1982. "Replicators and Vehicles". King's College Sociobiology Group, eds., *Current Problems in Sociobiology*. Cambridge University Press, Cambridge. pp. 45-64.

Dawkins, R. 1994. Burying the Vehicle. Peer commentary to Sober and Wilson 1994. *Behavior and Brain Sciences*. 17(4):616-617.

Dawkins, R. 2000. Obituary for W.D. Hamilton. *The Independent*. http://edge.org/conversation/w-d-hamilton-an-obituary.

Dugatkin, L.A. 2006. *The Altruism Equation: Seven Scientists Search for the Origins of Goodness*. Princeton University Press, Princeton.

Dugatkin, L.A., H.K. Reeve. 1994. Behavioral ecology and levels of selection: dissolving the group selection controversy. *Advances in the Study of Behavior*. 23:101-133.

Eldakar, O.T., and D.S. Wilson. 2011. Eight Criticisms Not to Make About Group Selection. *Evolution*. 65(6):1523-1526.

Fisher, R.A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*. 3:399–433

Fisher, R.A. 1937. The design of experiments. Oliver & Boyd, Edinburgh.

Fisher, R.A. 1941. Average excess and average effect of a gene substitution. *Annals of Eugenics*. 11:53–63

Fisher, R.A. 1953. Population genetics. The Croonian lecture. *Proceedings of the Royal Society B: Biological Sciences*. 141:510–523

Fisher, R.A. 1958. *The genetical theory of natural selection*. 2nd edition of 1930 Oxford University Press 1st edition. Dover Publications Inc., New York.

Frank, S. 2011. Pathology from Evolutionary Conflict. Presentation at Sackler Colloquia In Light of Evolution V: Cooperation. Irvine, California, January 7-8.

Garcia, C., and M.A. Toro. 1990. Individual and Group Selection for Productivity in *Tribolium castaneum*. *Theoretical and Applied Genetics*. 79(2):256-260.

Getty, T. 1999. What do experimental studies tell us about group selection in nature? *American Naturalist*. 154(5):596-598.

Ghiselin, M.T. 1974. A Radical Solution to the Species Problem. *Systematic Zoology*. 23(4): 536-544.

Gilbert, S.F., J. Sapp, and A.I. Tauber. 2013. A Symbiotic View of Life: We Have Never Been Individuals. *The Quarterly Review of Biology*. 87(4):325-341.

Goodnight, C. J. 1985. The influence of environmental variation on group and individual selection in a cress. *Evolution*. 39:545–558.

Goodnight, C.J. 1990a. Experimental studies of community evolution I: The response to selection at the community level. *Evolution*. 44(6):1614-1624.

Goodnight, C.J. 1990b. Experimental studies of community evolution II: The ecological basis of the response to community selection. *Evolution*. 44(6):1625-1636.

Goodnight, C.J. and L. Stevens. 1997. Experimental Studies of Group Selection: What do they tell us about group selection in nature?. *American Naturalist*. 150, Supplement.

Goodnight, C.J., J.M. Schwartz, L. Stevens. 1992. Contextual Analysis of Models of Group Selection, Soft Selection, Hard Selection, and the Evolution of Altruism. *The American Naturalist*. 140(5):743-761.

Gordon, J. N. Knowlton, D.A. Relman, F Rohwer, and M. Youle. 2013. Superorganisms and Holobionts. *Microbe*. 8(4)152-153.

Gould, S.J. 2002. *The Structure of Evolutionary Theory*. Belknap Press of Harvard University Press, Cambridge.

Grillenberger, B.K., T. Koevoets, M.N. BurtonChellew, E.M. Sykes, D.M. Shuker, L. Van de Zande, R. Bijlsma, J. Gadau, and L.W. Beukeboom. 2008. Genetic structure of natural *Nasonia vitripennis* populations: validating assumptions of sex-ratio theory. *Molecular Ecology*. 17:2854-2864.

Hamilton, A., and C.C. Dimond. 2012. Groups, individuals, and evolutionary restraints: the making of the contemporary debate over group selection. *Biology and Philosophy*. 27(2):299–312

Hamilton, W. D. 1967. Extraordinary sex ratios. *Science*. 156:477– 488.

Hamilton, W.D. 1964. "The Genetical Evolution of Social Behaviour" (I and II). *Journal of Theoretical Biology* 7: 17–52.

Hamilton, W.D. 1996. *Narrow Roads of Gene Land I*. Freeman, Oxford.

Harmon, O. 2010. *The Price of Altruism*. W.W. Norton and Company, New York.

Heisler, I. L., and J. Damuth. 1987. A method for analyzing selection in hierarchically structured populations. *The American Naturalist*. 130:582–602.

Henrich, Natalie, and Joseph Henrich. 2007. *Why Humans Cooperate: A cultural and evolutionary explanation*. Oxford University Press, Oxford.

Hodgson, G.M. 1991. Hayek Theory of Cultural Evolution – An Evaluation in the Light of Vanberg Critique. *Economics and Philosophy*. 7(1):67-82.

Hull, D. 2001. Interactors Versus Vehicles. In *Science and Selection: Essays on Biological Evolution and the Philosophy of Science*. Cambridge University Press, Cambridge.

Hull, D.L. 1976. Are Species Really Individuals?. *Systematic Zoology*. 25(2): 174-191

Hull, D.L. 1978. A Matter of Individuality. *Philosophy of Science*. 45(3): 335-360.

Hull, D.L. 1980. Individuality and Selection. *Annual Review of Ecology, Evolution, and Systematics*. 11:311-31

Huxley, T.H. 1893. "The Struggle for Existence in Human Society." In *Evolution and Ethics*, by T. H. Huxley. 9:195–236. Macmillan, London.

Ivens AB, D.M. Shuker, L.W. Beukeboom, I. Pen. 2009. Host acceptance and sex allocation of Nasonia wasps in response to conspecifics and heterospecifics. *Proceedings of the Royal Society B: Biological Sciences*. 276:3663-3669.

Joblonski, D. 2008. Species Selection: Theory and Data. *Annual Review of Ecology, Evolution, and Systematics*. 39:501-524.

Keller, E.F., and E.A. Lloyd. 1992. *Keywords in Evolutionary Biology*. Harvard University Press, Cambridge.

Kelley, J.K. 1997. Fitness variation across a subdivided population of the annual plant *Impatiens capensis*. *Evolution*. 51(4):1100-1111.

Kerr, B., G. Neuhauser, B.J.M. Bohannan, A.M. Dean. 2006. Local migration promotes competitive restraint in a host-pathogen "tragedy of the commons." *Nature*. 442:75-78.

Kropotkin, P.A. 1902. *Mutual Aid: A Factor in Evolution*. McClure Philips, New York.

Lack, D.L. 1954. *The natural regulation of animal numbers*. Oxford University Press, Oxford.

Lack, D.L. 1966. *Population Studies of Birds*. Clarendon Press, Oxford.

Leigh, E.G., Jr. 2010. The Group Selection Controversy. *Journal of Evolution Biology*. 23:6-19.

Lewontin, R. C. 1970. The Units of Selection. *Annual Review of Ecology and Systematics*. Vol 1:1-18

Lewontin, R.C. 1962. Interdeme selection controlling a polymorphism in the house mouse. *The American Naturalist*. 96: 65–78.

Lieberman, B.S. and E.S. Verba. 1995. Hierarchy Theory, Selection, and Sorting – A Phylogenetic Perspective. *Bioscience*. 45(6) 394-399.

Lloyd E.A. 2005. Units and Levels of Selection. The Stanford Encyclopedia of Philosophy (Fall 2005 Edition), E.N. Zalta (ed.) URL: <http://plato.stanford.edu/archives/fall2005/entries/selection-units/>.

Lloyd, E.A. 2001. "Units and Levels of Selection: An Anatomy of the Units of Selection Debates." In: *Thinking about Evolution: Historical Philosophical and Political Perspectives*. 276-290.

Matessi, C., Jayakar, S.D. 1973. A model for the evolution of altruistic behavior. *Genetics*. 74:S174.

Matessi, C., Jayakar, S.D. 1976. Conditions for the evolution of altruism under Darwinian selection. *Theoretical Population Biology*. 9:360-87.

Maynard Smith, J. 1964. Group selection and kin selection. *Nature*. 201:1145-47.

Maynard Smith, J. 1998. The Units of Selection. *Novartis Foundation Symposium*. 213:203-217.

Maynard Smith, J. and E. Szathmary. 1995. *The Major Transitions in Evolution*. Oxford University Press, Oxford.

Maynard Smith, John. 1997. "The first idea of population regulation by group selection." Online video. www.webofstories.com/play/john.maynard.smith/. Posted 2008. Accessed 2/3/2014.

McCauley, D.E. and M.J. Wade. 1980. Group Selection – The Genetic and Demographic Basis for the Phenotypic Differentiation of Small Populations of *Tribolium castaneum*. *Evolution*. 34(4):812-821.

McFall-Ngaia, M., M.G. Hadfieldb, T.C.G. Bosch, H.V. Carey, T. Domazet-Loso, A.E. Douglas, N. Dubilier, G. Eberl, T. Fukami, S.F. Gilbert, U. Hentschel, N. King, S. Kjelleberg, A.H. Knoll, N. Kremer, S.K. Mazmanian, J.L. Metcalf, K. Nealson, N.E. Pierce, J.F. Rawls, A. Reid, E.G. Ruby, M. Rumpho, J.G. Sanders, D. Tautz, and J.J. Wernegreen. 2013. Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Science*. 110(9):3229-3236.

Michod, R.E., and A.M. Nedelcu. 2003. On the Reorganization of Fitness During Evolutionary Transitions in Individuality. *Integrative and Comparative Biology*. 43:64–73.

Miller, J.H. and S.E. Page. 2007. *Complex Adaptive Systems*. Princeton University Press, New Jersey.

Molbo D., and E.D. Parker Jr. 1996. Mating structure and sex ratio variation in a natural population of *Nasonia vitripennis*. *Proceedings of the Royal Society of London — Series B: Biological Sciences*. 263:1703 – 1709.

Muir, W. M. 1996. Group selection for adaptation to multiple-hen cages: selection program and direct responses. *Poultry Science*. 75:447–458.

Nowak, M. A., C. E. Tarnita, and E. O. Wilson. 2010. The evolution of eusociality. *Nature*. 466:1057–1062.

Nowak, M. A., C. E. Tarnita, and E. O. Wilson. 2011. Nowak *et al*. reply. *Nature*. 471:E9–E10

Okasha S. 2004a. The ‘‘averaging fallacy’’ and the levels of selection. *Biology and Philosophy*. 19:167–184

Okasha, S. 2004b. Multilevel selection and the partitioning of covariance: a comparison of three approaches. *Evolution*. 58:486–494.

Okasha, S. 2006. *Evolution and the levels of selection*. Oxford University Press, Oxford.

Orzack, S.H., E.D. Parker Jr., and J. Gladstone. 1991. The Comparative Biology of Genetic Variation for Conditional Sex Ratio Behavior in a Parasitic Wasp, *Nasonia vitripennis*. *Genetics*. 127:583-599.

Page R.E. Jr., and G.V. Amdam. 2007. The Making of A Social Insect: Developmental Architectures of Social Design. *Bioessays*. 29(4):334-343.

Page, R.E. Jr., and M.K. Fondrk. 1995. The Effects of colony-level selection on the social organization of honey been (Aspis melifera L.) colonies: colony-level components of pollen hoarding. *Behavioral Ecology and Sociobiology*. 36:135-144.

Parker, D.E., and S.H. Orzack. 1985. Genetic Variation for the Sex Ratio in *Nasonia vitripennis*. *Genetics*. 110:93-105.

Price, G. R. 1970. Selection and covariance. *Nature*. 227:520-21.

Price, G. R. 1972. Extension of covariance selection mathematics. *Annals of Human Genetics*. 35:485-90.

Racey, P. 1997. Obituary: Vero Wynne-Edwards. *The Independent*. London, England. http://www.royalsoced.org.uk/cms/files/fellows/obits_alpha/wynne-edwards_vero.pdf

Ratclif, W.C., J.T. Pentz, M. Travisano. 2013. Tempo and Mode of Multicellular Adaptation In Experimentally Evolved *Saccharomyces cerevisiae*. *Evolution*. 7(6):1573–1581.

Reeve, H.K., and L. Keller. 1999. Levels of Selection: Burying the Units-of-Selection debate and Unearthing the Crucial New Issues. 3-14. In *Levels of Selection in Evolution*, edited by L. Keller. Princeton University Press, New Jersey.

Richerson, P.J. and R. Boyd. 2005. *Not by Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press. Chicago.

Rosenberg, E., O. Koren, L. Reshef, R. Efrony and I. Zilber-Rosenberg. 2007. The role of microorganisms in coral health, disease and evolution. *Nature Reviews*. 5:344-362.

Singh, Y., J. Ahmad, J. Musarrat, N. Z. Ehtesham, and S.E. Hasnain. 2013. Emerging importance of holobionts in evolution and in probiotics. *Gut Pathogens*. 5(12):1-8.

Slatkin, M., and M.J. Wade. 1978. Group Selection on a Quantative Character. *Proceedings of the National Academy of Science*. 75(7):3531-3534.

Sober, E. 1984. Sets, species, and evolution: comments on Philip Kitcher's ''species''. Journal of the Philosophy of Science Association. 51:334–341.

Sober, E. and D.S. Wilson. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Harvard University Press, Cambridge.

Sober, E., and D.S. Wilson. 1994. A Critical Review of Philosophical Work on the Units of Selection Problem. *Philosophy of Science*. 61(4):534-555.

Sterns, S.C. 1982. The Emergence of Evolutionary and Community Ecology as Experimental Sciences. *Perspectives in biology and Medicine*. 25(4):621-648.

Stevens, L., C.J. Goodnight, and S. Kalisz. 1995. Multilevel Selection in Natural Populations of Impatiens capensis. *The American Naturalist*. 145(4):513-526).

Sturtevant, A.H. 1938. Essays on evolution II. On the effects of selection on social insects. *Quarterly Review of Biology*. 13:74–76.

Swenson, W., D.S. Wilson, and R. Elias. 2000a. Artificial Ecosystem Selection. *Proceedings of the National Academy of Science*. 97(16):9110-9114.

Swenson, W., J. Arendt, and D.S. Wilson. 2000b. Artificial Selection of microbial Ecosystems for 3-chloroaniline biodegradation. *Environmental Microbiology*. 2(5):564-571.

Szathmary, E., and J. Maynard Smith. 1995. The major evolutionary transitions. *Nature*. 374:227-232.

Treisman, M. 1983, Errors in the theory of the structured deme. *Journal of Theoretical Biology*. 102(3):339-346.

Vrba, E.S. 1984. What is Species Selection? *Systematic Zoology*. 33(3):318-328.

Wade, M.J. 1976. Group Selection among Laboratory Populations of *Tribolium*. *Proceedings of the National Academy of Sciences*. 73:4604–4607.

Wade, M.J. 1977. An experimental study of group selection. *Evolution*. 31:134–153.

Wade, M.J. 1978. A Critical Review of the Models of Group Selection. *The Quarterly Review of Biology*. 53(2): 101-114

Wade, M.J. 1979. Primary Characteristics of *Tribolium* Populations Group Selected for Increased and Decreased Population-Size. *Evolution*. 33(2)749-764.

Wade, M.J. 1980. Group Selection, Population Growth Rate, and Competitive Ability in the Flour Beetles *Tribolium* spp. *Ecology*. 61(5)1056-1064.

Wade, M.J. 1982. Group Selection – Migration and the Differentiation of Small Populations. *Evolution*. 36(5)949-961.

Wade, M.J. and C.J. Goodnight. 1991. Wright Shifting Balance Theory – An Experimental Study. *Science*. 253(5023):1015-1018.

Wade, M.J. and D.E. McCauley. 1980. Group Selection – The Phenotypic and Genotype Differentiation of Small Populations. *Evolution*. 34(4):799-812.

Wade, M.J., D.S. Wilson, C. Goodnight, D. Taylor, Y. Bar-Yam, M.A.M. de Aguiar, B. Stacey, J. Werfel, G.A. Hoelzer, E.D. Brodie III, P. Fields, F. Breden, T.A. Linksvayer, J.A. Fletcher, P.J. Richerson, J.D. Bever, J.D. Van Dyken, and P. Zee1. 2010. Multilevel and kin selection in a connected world. *Nature*. 463: E8-E9.

Werfel, J., and Y. Bar-Yam. 2004. The Evolution of Reproductive Restraint through Social Communication. *Proceedings of the National Academy of Sciences*. 101 (30): 11019–11024.

Werren, J.H. 1980. Sex Ratio Adaptations to Local Mate Competition in a Parasitic Wasp. *Science*. 208(4448):1157-1159.

Werren, J.H., S. Richards, C.A. Desjardins, O. Niehuis, J. Gadau, J.K. Colbourne, *et al*. 2010. Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. *Science*: 327(5963)343 - 348.

West, S.A., A.S. Griffin, A. Gardner. 2007. Social semantics: altruism, cooperation, mutualism, strong reciprocity, and group selection. *Journal of Evolutionary Biology*. 20(2):415-432.

West, S.A., A.S. Griffin, and A. Gardner. 2008. Social semantics: how useful has group selection been? *Journal of Evolutionary Biology*. 21:374–385.

Whitham, T.G., J.K. Bailey, J.A. Schweitzer, S.M. Shuster, R.K. Bangert, C.J. LeRoy, E.V. Lonsdorf, G.J. Allan, S.P. DiFazio, B.M. Potts, D.G. Fischer, C.A. Gehring, R.L. Lindroth, J.C. Marks, S.C. Hart, G.M. Wimp, S.C. Wooley. 2006. A Framework for Community and Ecosystem Genetics: From Genes to Ecosystems. *Nature Reviews*. 7:510-523.

Whitham, T.G., W.P Young, G.D. Martinsen, C.A. Gehring, J.A. Schweitzer, S.M. Schuster, G.M. Wimp, D.G. Fischer, J. K. Bailey, R.L. Lindroth, S. Woolbright, and C.R. Kuske. 2003. Community and Ecosystem Genetics: A Consequence of the Extended Phenotype. *Ecology*. 84(3):559-573.

Wild, G., A. Gardener, and S.A. West. 2010. Wild, Gardner & West reply. *Nature*. 463: E9-E10.

Wild, G., A. Gardner, S.A. West. 2009. Adaptation and the evolution of parasite virulence in a connected world. *Nature*. 459:983–986.

Williams, G.C. 1966. *Adaptation and Natural Selection*. Princeton University Press, Princeton.

Williams, G.C. 1992. *Natural selection: domains, levels, and challenges*. Oxford University Press, Oxford.

Williams, G.C. 1996. Preface (1996), in *Adaptation and Natural Selection*. Princeton University Press. Princeton.

Wilson, D. S. 1975. A theory of group selection. *Proceedings of the National Academy of Sciences*. 72:143-46

Wilson, D.S. 1983. The Group Selection Controversy: History and Current Status. *Annual Review of Ecology and Systematics*. 14: 159-187

Wilson, D.S. 1997. Biological Communities as Functionally Organized Units. *Ecology*. 78(7):2018-2024.

Wilson, D.S. 2008. Social Semantics: Toward a Genuine Pluralism in the Study of Social Behavior. *Journal of Evolutionary Biology* 21:368–373.

Wilson, D.S. 2008. Social semantics: toward a genuine pluralism in the study of social behaviour. *Journal of Evolutionary Biology*. 21:368–373.

Wilson, D.S., and E. Sober. 1994. Reintroducing Group Selection to the Human Behavioral Sciences. *Behavioral and Brain Sciences*. 17(4):585-608.

Wilson, D.S., and E.O. Wilson. 2007. Rethinking the Theoretical Foundation of Sociobiology. *The Quarterly Review of Biology*. 82(4):327-348.

Wilson, DS. 1997. Biological communities as functionally organized units. *Ecology*. 78(7):2018-2024.

Winther, R.G., M.J. Wade, and C.C. Dimond. 2013. Pluralism in evolutionary controversies: styles and averaging strategies in hierarchical selection theories. *Biology and Philosophy*. 28(6):957-979

Wright, S. 1929. Evolution in a Mendelian population. *The Anatomical Record*. 44:287.

Wright, S. 1932. *The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution*. Cornell University Press, Ithaca.

Wright, S. 1945. Tempo and mode in evolution: a critical review. *Ecology*. 26:415–419.

Wright, S. 1978. The relation of livestock breeding to theories of evolution. *Journal of Animal Science*. 46: 1192–1200.

Wright, S. 1980. Genic and organismic selection. *Evolution*. 34:825–843.

Wynne-Edwards, V.C. 1962. *Animal Dispersion in Relation to Social Behavior*. Oliver and Boyd, Edinburgh.

Wynne-Edwards, V.C. 1963. Intergroup Selection in the Evolution of Social Systems. *Nature*. 200(4907):623-626.

Wynne-Edwards, V.C. 1964. Group Selection and Kin Selection: Response to Maynard-Smith. *Nature*. 201(4924):1145–1147.

Wynne-Edwards, V.C. 1985. "Backstage and upstage with "animal dispersion"". In: Dewsbury D (ed) *Studying animal behavior: autobiographies of the founders*. University of Chicago Press, Chicago

Wynne-Edwards, V.C. 1993. A Rationale for Group Selection. *Journal of Theoretical Biology*. 162:1-22.

APPENDIX A

SUPPLEMENTAL MATERIALS FOR CHAPTER 1

Source Strains for ASURx

The *Wolbachia* that infect *N*. *vitripennis* come from two major stains. Even within the species, there can be cytoplasmic incompatibilities when *N*. *vitripennis* individuals carrying different strains mate (Bordenstein and Werren 2007). To avoid incompatibilities without resorting to antibiotics, the TMPRx strain was created by crossing eight strains of *N*. *vitripennis* against the LabII strain, which is known to posses both major strains of *N*. *vitripennis Wolbachia*. All nine strains used, the 8 strains plus the LabII strain, were received from the Werren laboratory at the University of Rochester during spring 2013 (Fig. S1).

Creating the ASURx

Approximately 10 males from each of the 8 strains recieved from the Werren laberatory were placed in a vial with 10 virgin females of the LabII strain. This created 8 strains of *N*. *vitripennis* with the nuclear DNA that was standardized against the LabII cytoplasm and *Wolbachia*. After 1 generation, the 8 resulting standardized strains were combined into a single TMPRx population spread across 4 rearing tubes (Fig. S1). This was done by taking 8 mated females from each standardized strain, and placing 2 into each of 4 rearing tubes (Fig. S1). Thus, the experimental population was formed by 2 individuals from each of 8 strains for a total of 16 females in each tube, with 4 tubes for a total population of 64.

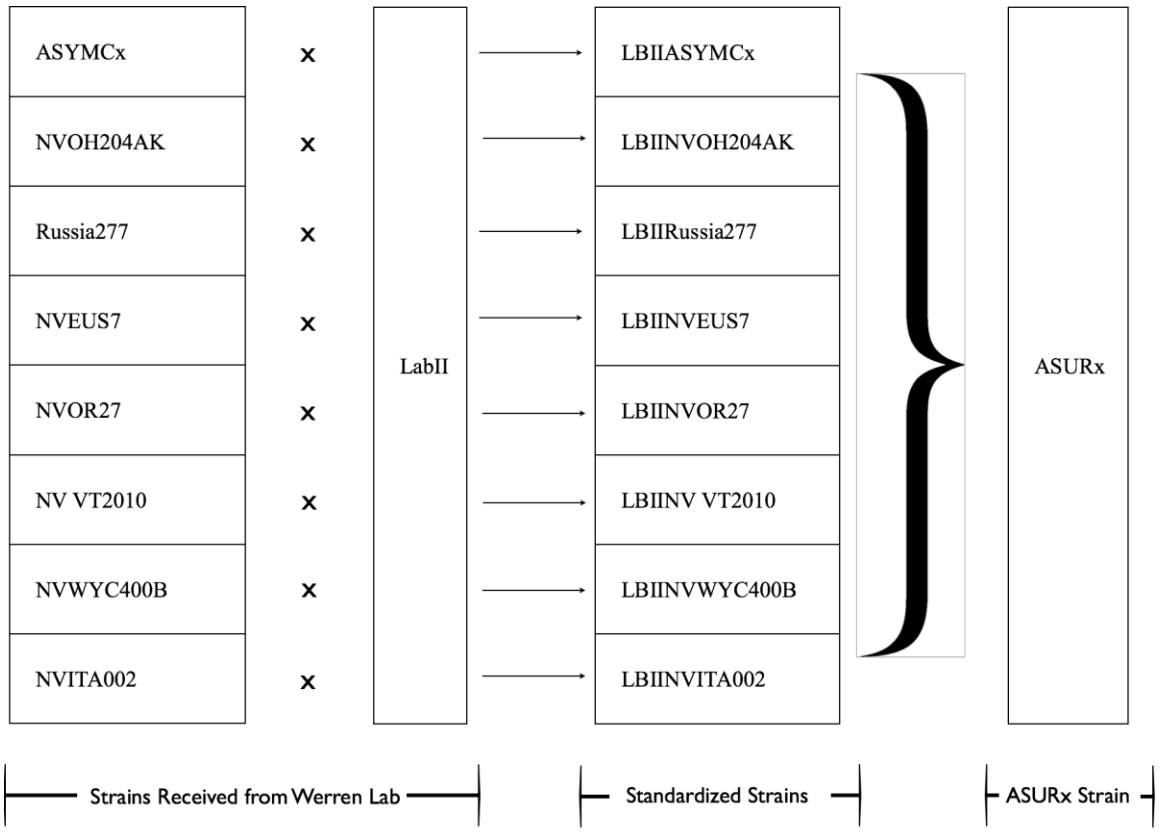


Figure S.1 Creation of the ASURx laboratory population**.**

Maintenance of ASURx

To maintain allelic diversity in this population, I used the remixing techniques of Pannebakker and van de Zande (personal correspondence) for maintaining their High Variation Remix (HVRx) population of *N*. *vitripennis*. Each generation, after the adults emerged and mated, approximately half of the population was transferred to a new rearing tube. Then 40 host pupae were placed into each of the four new rearing tubes. After seven days, the hosts from each tube were collected and evenly redistributed across the four tubes. In this way, the 40 hosts that had been in tube 1 were redistributed so that 10 were in tubes 1, 2, 3, and 4. Thus, after the redistribution, tube 1 consisted of 10 hosts from what had been tube 1, 10 hosts from what had been tube 2, 10 hosts from what had

been tube 3, and 10 hosts from what had been tube 4. After another seven days, the wasps emerge and were allowed to mate before moving the females to a new tube and beginning again.

Methods for transferring and counting *Nasonia*

The vials to be counted were placed in a bucket of icewater, then emptied onto a chilled petri dish left in contact with the icewater. This slowed the *Nasonia*, allowing for easy manipulation. After the population was counted, wasps were placed back in the vial and allowed to sit at room temperature until all adults were again active. The selected vial was then sealed with a cap that had a small hole so that mated females could move out of the vial and into another vial placed on top. The first five females that did so in the group treatment were used to start the first collective for the next generation, and four collective vials were subsequently created in the same way. Within the organismal treatment, five females were collected in the same way but each female was placed into a solitary vial to create the next generation.